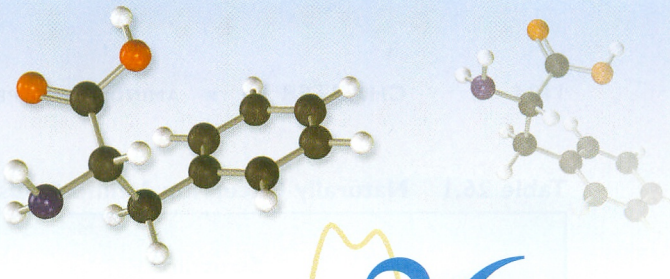
VISIT…



Amino Acids, Peptides, and Proteins

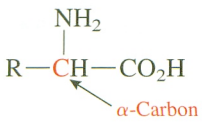
CHAPTER 26

AMINO ACIDS AND THE polymers that are formed from them, peptides and proteins, are among the most important compounds found in living organisms. They act as structural components, catalysts, hormones, neurotransmitters, and essential nutrients, among other things. This chapter emphasizes the organic chemistry of these important biomolecules. Details of their synthesis and function in living organisms are left for biochemistry textbooks and are only briefly described here.

First the general structure and chemistry of the amino acids is presented. Then several methods that can be used to prepare them in the laboratory are discussed. After an introduction to the structure of peptides and proteins, chemical methods that can be used to determine the amino acid sequence in proteins are presented. Next, the synthesis of peptides in the laboratory is introduced. Finally, the three-dimensional structure of proteins and the mechanism of action of enzymes are briefly addressed.

26.1 AMINO ACIDS

As the name implies, amino acids are carboxylic acids that contain amino groups. The amino acids that are of most importance in nature have the amino group on carbon 2 of the carboxylic acid (the α-carbon), as shown in the following general structure, and are therefore sometimes called α-amino acids:



An α-amino acid

Twenty "standard" amino acids commonly occur in nature. They differ in the structure of the side chain that is attached to the α-carbon (the R group in the previous structure). Table 26.1 shows the structures of the standard amino acids at pH 7, along with their names and abbreviations.

MASTERING ORGANIC CHEMISTRY

- Recognizing the General Structure, Including Stereochemistry, of Amino Acids
- Understanding the Acid–Base Chemistry of Amino Acids
- Predicting the Products of the General Chemical Reactions of Amino Acids
- Using Organic Reactions to Synthesize Amino Acids
- Understanding How Peptides Are Sequenced
- Understanding the Laboratory Synthesis of a Peptide

ORGANIC Chemistry Now™

Look for this logo in the chapter and go to OrganicChemistryNow at **http://now.brookscole.com/hornback2** for tutorials, simulations, problems, and molecular models.

Table 26.1 Naturally Occurring Amino Acids

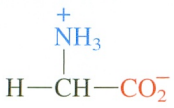
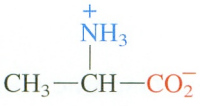
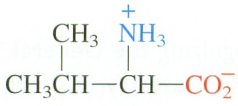
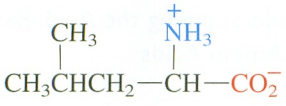
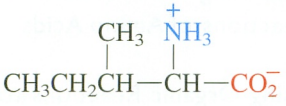
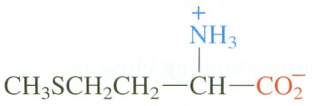
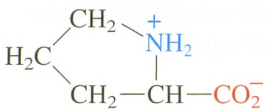
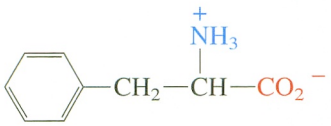
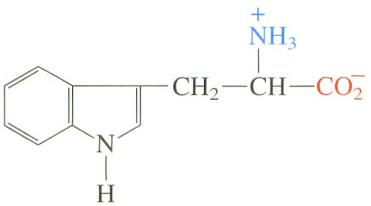
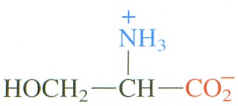
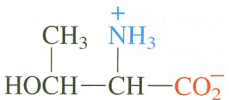
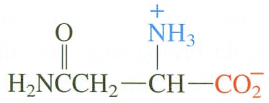
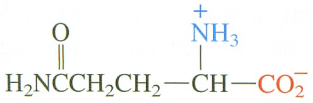
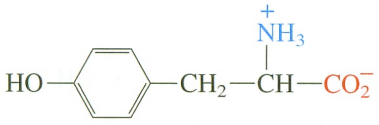
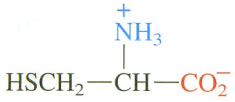
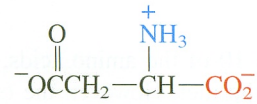
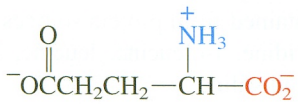
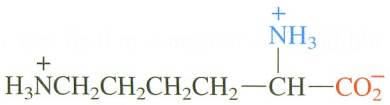
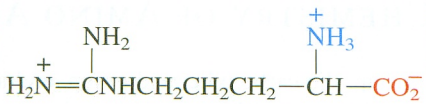
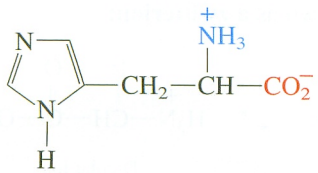
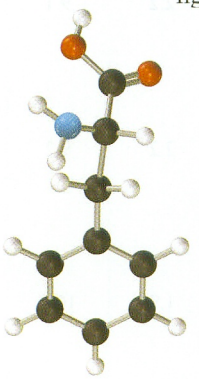
	Structure	Name	Abbreviation	pK_a of α-CO_2H	pK_a of α-NH_3^+	pK_a of R Group	pI
Nonpolar Side Chains		Glycine	Gly or G	2.3	9.8		6.1
		Alanine	Ala or A	2.3	9.9		6.1
		Valine	Val or V	2.3	9.7		6.0
		Leucine	Leu or L	2.3	9.7		6.0
		Isoleucine	Ile or I	2.3	9.7		6.0
		Methionine	Met or M	2.1	9.3		5.7
		Proline	Pro or P	2.0	10.6		6.3
		Phenylalanine	Phe or F	2.2	9.3		5.7
		Tryptophan	Trp or W	2.5	9.4		5.9
Polar Side Chains		Serine	Ser or S	2.2	9.2		5.7

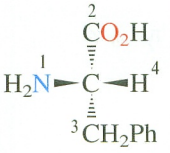
Table 26.1 Naturally Occurring Amino Acids—cont'd

	Structure	Name	Abbreviation	pK_a of α-CO_2H	pK_a of α-NH_3^+	pK_a of R Group	pI
Polar Side Chains—cont'd		Threonine	Thr or T	2.1	9.1		5.6
		Asparagine	Asn or N	2.1	8.7		5.4
		Glutamine	Gln or Q	2.2	9.1		5.7
		Tyrosine	Tyr or Y	2.2	9.2	10.5	5.7
		Cysteine	Cys or C	1.9	10.7	8.4	5.2
Acidic Side Chains		Aspartic acid	Asp or D	2.0	9.9	3.9	3.0
		Glutamic acid	Glu or E	2.1	9.5	4.1	3.1
Basic Side Chains		Lysine	Lys or K	2.2	9.1	10.5	9.8
		Arginine	Arg or R	1.8	9.0	12.5	10.8
		Histidine	His or H	1.8	9.3	6.0	7.6

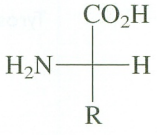
In this table the compounds have been grouped according to the polarity of their side chains (nonpolar, polar, acidic, and basic).

All of the amino acids, except the simplest one, glycine (R = H), have four different groups attached to the α-carbon and are, therefore, chiral. In general, only a single enantiomer is found in nature. When the Fischer projection formula for the naturally occurring enantiomer of an amino acid is drawn in the conventional manner, with the carboxylic acid group at the top and the carbon chain vertical, the amino group is on the left. Because D-glyceraldehyde has its hydroxy group on the right, the naturally occurring amino acids belong to the L series. For most amino acids the NH_2 group has the highest priority according to the Cahn-Ingold-Prelog sequence rules, followed in order by the CO_2H group, the R group, and the H. These amino acids have the *S* absolute configuration. However, in the case of cysteine the CH_2SH group has a higher priority than the CO_2H group; thus, the absolute configuration of cysteine is *R*. Note that several of the amino acids also contain a stereocenter in their side chains. Again, only one configuration at this carbon is usually found.

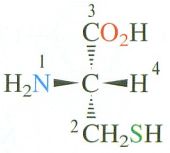




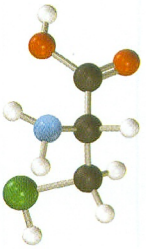
L-Phenylalanine
(*S*)-phenylalanine



L-Amino acid



L-Cysteine
(*R*)-cysteine



Human beings are able to synthesize 10 of the amino acids, termed *nonessential amino acids,* from other compounds in the diet. However, the other 10 amino acids, termed **essential amino acids,** cannot be synthesized by humans (or are synthesized only in small amounts) and must be obtained from protein sources in the diet. The essential amino acids are arginine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine.

PROBLEM 26.1

Which of the amino acids have an additional stereocenter in their side chains?

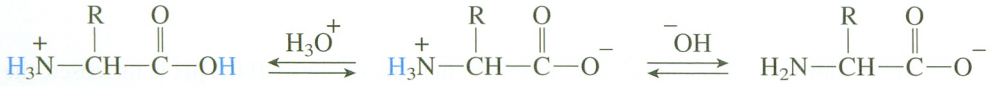
26.2 Acid–Base Chemistry of Amino Acids

Amino acids have both an acidic group, the carboxylic acid group, and a basic group, the amino group. These two functional groups undergo an intramolecular acid base reaction to form a **dipolar ion,** also known as a **zwitterion:**

Dipolar ion
or
zwitterion

A carboxylic acid is a stronger acid ($pK_a \approx 5$) than the conjugate acid of an amine ($pK_a \approx 9$). Therefore, the equilibrium in the preceding reaction favors the dipolar ion, and this is the species that is present in the solid form of an amino acid. Thus, many of the physical properties of amino acids resemble those of salts rather than those of typical organic compounds. They have high melting points (in the vicinity of 300°C), are quite soluble in water, and are rather insoluble in typical organic solvents. Although amino acids exist as dipolar ions, they are often written with their amino and carboxylic acid groups in un-ionized form.

The dipolar ion has both an acidic group, the ammonium cation, and a basic group, the carboxylate anion, so it can act as either an acid or a base. Such compounds are termed **amphoteric.** The species that is present in aqueous solution depends on the pH. In the pH range near neutral the amino acid is present in the form of the dipolar ion. In acidic solution the carboxylate group becomes protonated and the amino acid is present as a cation, whereas in basic solution the ammonium group gives up a proton and the molecule exists as an anion:



Cation, pH < 2 — Dipolar ion, pH ≅ 6 — Anion, pH > 10

The concentration of an acid and its conjugate base in aqueous solution can be calculated from its K_a or pK_a. Recall from Chapter 4 that

$$K_a = \frac{[H_3O^+][A^-]}{[HA]} \quad \text{and} \quad pK_a = -\log K_a$$

Therefore,

$$pK_a = -\log \frac{[H_3O^+][A^-]}{[HA]} = -\log[H_3O^+] - \log \frac{[A^-]}{[HA]}$$

and

$$pK_a = pH - \log \frac{[A^-]}{[HA]}$$

or

$$pH = pK_a + \log \frac{[A^-]}{[HA]} \quad \text{(Henderson-Hasselbach equation)}$$

From the Henderson-Hasselbalch equation, when $pH = pK_a$, then

$$\log \frac{[A^-]}{[HA]} = 0$$

and

$$[HA] = [A^-]$$

Let's use the Henderson-Hasselbalch equation to examine the effect of pH on the concentrations of the various forms of glycine:

Cationic form, $pK_{a1} = 2.3$ — Dipolar ion, $pK_{a2} = 9.8$ — Anionic form

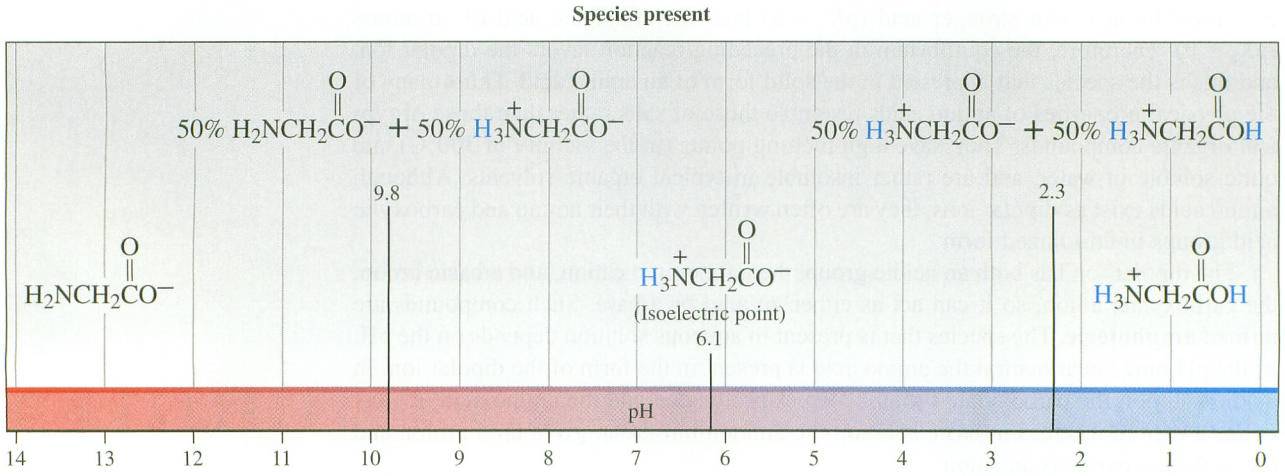


Figure 26.1

FORMS OF GLYCINE PRESENT AT VARIOUS pHs.

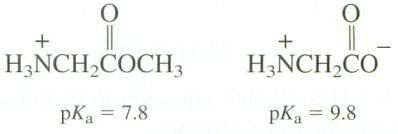
At very low pH, glycine is almost entirely in the cationic form. If the pH is increased, the cationic form gives up a proton from the carboxylic acid group (the carboxylic acid is a stronger acid than the ammonium group), producing the dipolar ion. From the Henderson-Hasselbalch equation the concentration of the cationic form equals the concentration of the dipolar ion when the pH is equal to pK_{a1}—that is, at pH = 2.3. If the pH is increased further, the concentration of the dipolar ion increases until nearly all of the glycine is present in that form. Then the ammonium group begins to give up its acidic proton, producing the anionic form. The concentration of the dipolar ion equals that of the anionic form when the pH is equal to pK_{a2}—that is, at pH = 9.8. At higher pH, glycine is present predominantly in its anionic form. The concentration of the dipolar ion is at a maximum at a pH equal to the average of pK_{a1} and pK_{a2}, or pH = $\frac{1}{2}(2.3 + 9.8) = 6.1$ in the case of glycine. This pH is called the **isoelectric point** (pI) because the amino acid has an overall charge of zero; that is, it is neutral. The forms of glycine that are present at various pHs are illustrated in Figure 26.1.

PROBLEM 26.2

Explain why the carboxylic acid group of the cationic form of glycine ($pK_a = 2.3$) is a stronger acid than the carboxylic acid group of acetic acid ($pK_a = 4.8$).

PROBLEM 26.3

Explain the order of these pK_a values:

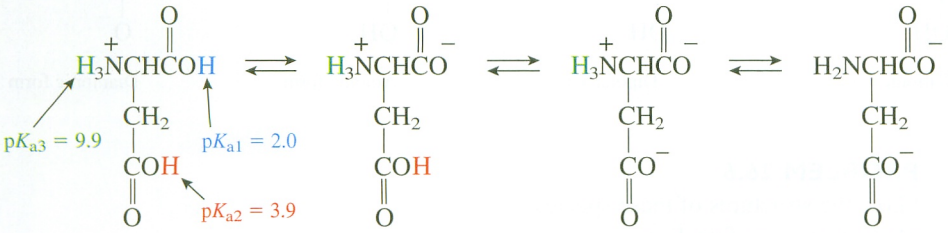


Two important points should be learned from this discussion. First, glycine (and other amino acids) is never present in aqueous solution in a neutral form with uncharged carboxylic acid and amino groups. It is present as a cationic form, a dipolar ion,

or an anionic form, depending on the pH. Second, because the pH of most physiological solutions is near 7, which is close to the isoelectric point of glycine, it is commonly present as a dipolar ion in biological fluids.

Table 26.1 lists the pK_a's of the amino acids along with their isoelectric points (pI). Both those with nonpolar side chains and those with polar side chains have pK_a values near those of glycine. Therefore, these all have isoelectric points near 6 and exist predominantly as dipolar ions at neutral pH.

However, some amino acids have an acidic or basic functional group in their side chains. In these cases there is another acidity constant to consider. Let's examine the case of aspartic acid, which has two carboxylic acid groups and an amino group. At low pH, aspartic acid is present in a cationic form:



As the pH is increased, the proton is removed first from the carboxylic acid group closest to the ammonium group to generate the dipolar ion, then from the other carboxylic acid group, to generate an anionic form, and finally from the ammonium group to generate a dianionic form. The pI is the average of pK_{a1} and pK_{a2} and equals 3.0. The concentration of the anionic form is a maximum at a pH of 6.9, the average of pK_{a2} and pK_{a3}. Therefore, at neutral pH, aspartic acid is present predominantly in its anionic form. The situation is quite similar for glutamic acid.

The amino acids with basic side chains—lysine, arginine, and histidine—all have pI values greater than 7. In strongly acidic solution they exist in dicationic forms, and all have significant amounts of a cationic form and a dipolar ion present at neutral pH.

PROBLEM 26.4

Explain why the carboxylic acid group in the side chain of aspartic acid ($pK_a = 3.9$) is a weaker acid than the main carboxylic acid group of the amino acid ($pK_a = 2.0$).

PROBLEM 26.5

Explain why the carboxylic acid group in the side chain of glutamic acid is a weaker acid than the carboxylic acid group in the side chain of aspartic acid.

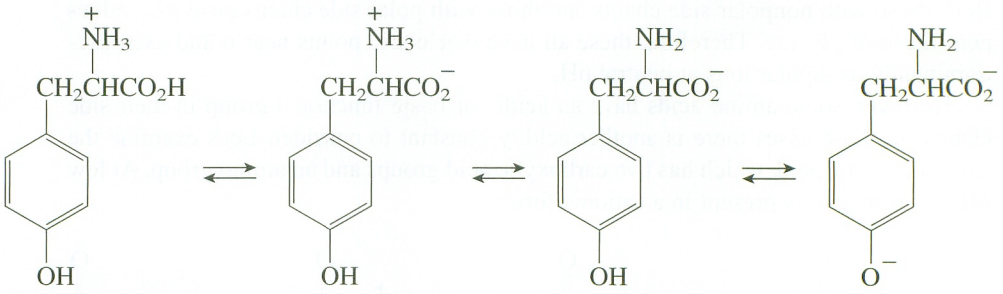
PRACTICE PROBLEM 26.1

Show the structure of the dianion form of tyrosine.

Solution

The pK_a's of the various groups of tyrosine are listed in Table 26.1. Starting from the cationic form that is present in strong acid, first a proton is removed from the carboxylic

acid group ($pK_a = 2.2$) to form the dipolar ion. The next proton is removed from the α-NH_3^+ ($pK_a = 9.2$) to give the anionic form. The final proton is removed from the OH group of the side chain ($pK_a = 10.5$) to give the dianionic form:



PROBLEM 26.6

Show the structures of these species:

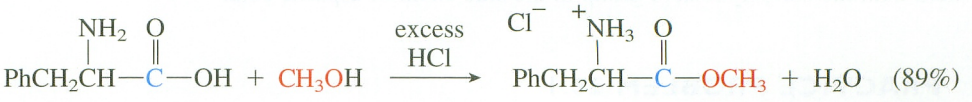
a) Dipolar ion of proline
b) Anion form of cysteine (careful)
c) Cation form of arginine

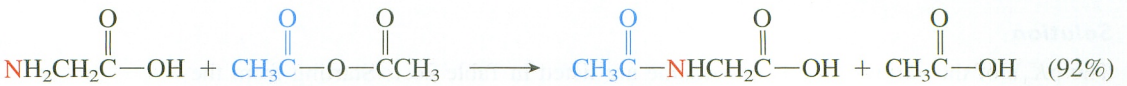
PROBLEM 26.7

a) Show the four differently charged forms of lysine.
b) Construct a diagram like that of Figure 26.1 for lysine.

26.3 Chemical Reactions of Amino Acids

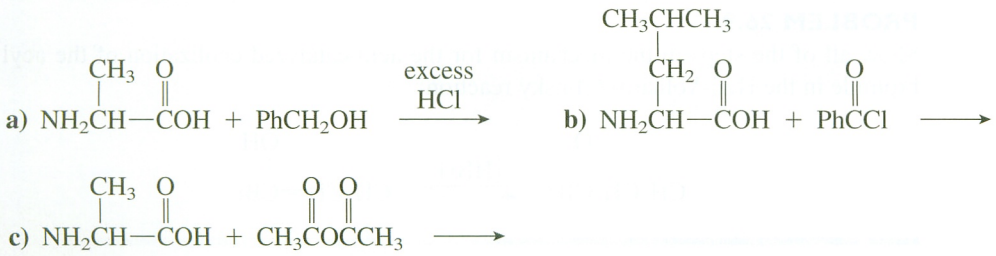
Amino acids exhibit chemical reactions that are typical of both amines and carboxylic acids. For example, the acid can be converted to an ester by the Fischer method. This reaction requires the use of an excess of acid because one equivalent is needed to react with the amino group of the product. As another example, the amine can be converted to an amide by reaction with acetic anhydride. Additional examples are provided by the reactions that are used in the preparation of peptides from amino acids described in Section 26.7:





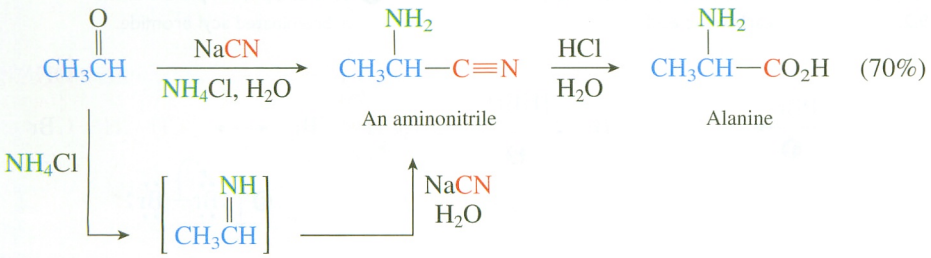
PROBLEM 26.8

Show the products of these reactions:



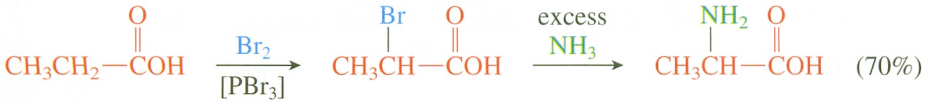
26.4 Laboratory Synthesis of Amino Acids

Because of their importance, a number of laboratory methods for the synthesis of amino acids have been developed. In the **Strecker synthesis** an aldehyde is treated with NaCN and NH_4Cl to form an aminonitrile, which is then hydrolyzed to the amino acid:



This reaction proceeds by initial reaction of ammonium chloride with the aldehyde to form an imine (see Section 18.8). Then cyanide adds to the imine in a reaction that is exactly analogous to the addition of cyanide to an aldehyde to form a cyanohydrin (see Section 18.4). The final step in the Strecker synthesis is the hydrolysis of the nitrile to a carboxylic acid (see Section 19.5).

In another method the amino group is introduced onto the α-carbon of a carboxylic acid. To accomplish this, an H on the α-carbon is first replaced with a Br, which can then act as a leaving group. The Br is replaced with an amino group by an S_N2 reaction with NH_3 as the nucleophile:



The bromine is introduced onto the α-carbon by treating the carboxylic acid with Br_2 and a catalytic amount of PBr_3 in a process known as the **Hell-Volhard-Zelinsky reaction.** This reaction proceeds through an enol intermediate. Because carboxylic acids form enols only with difficulty, a catalytic amount of PBr_3 is added to form a small amount of the acyl bromide, which enolizes more readily than the acid. Addition of bromine to the enol produces an α-bromoacyl bromide (see Section 20.2). This reacts with a molecule of the carboxylic acid in a process that exchanges the Br and OH groups to form the

product and another molecule of acyl bromide, which can go through the same cycle. Part of the mechanism for this process is outlined in Figure 26.2.

PROBLEM 26.9

Show all of the steps in the mechanism for the acid-catalyzed enolization of the acyl bromide in the Hell-Volhard-Zelinsky reaction:

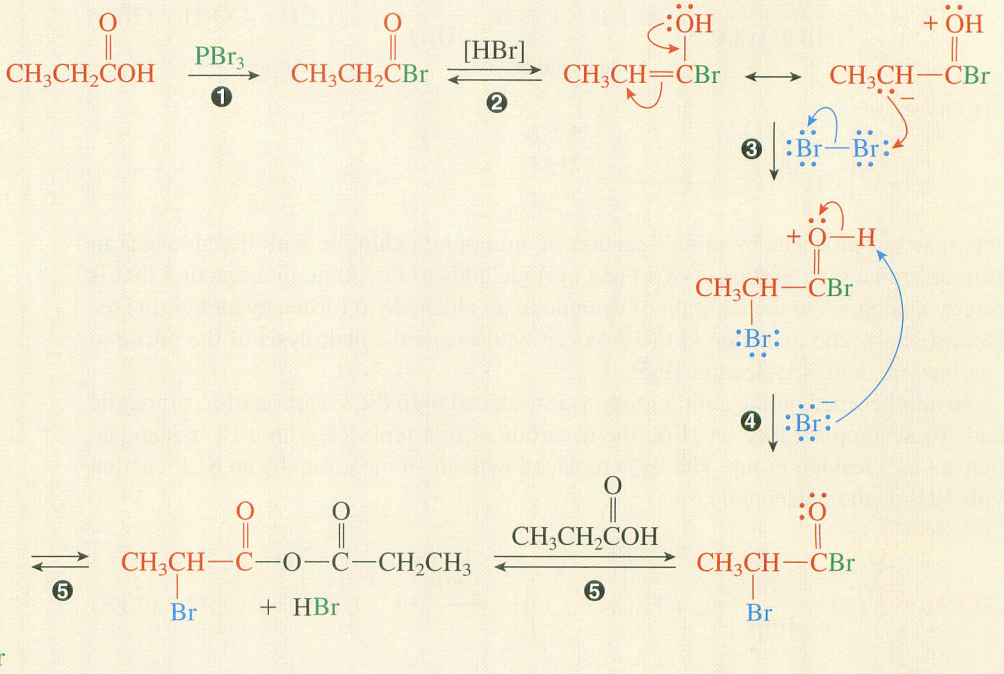
❶ The carboxylic acid reacts with PBr_3 to form an acyl bromide. The mechanism for this part of the reaction is very similar to that described for the formation of an acyl chloride in Section 19.2.

Next the acyl bromide is brominated at the α-carbon. This acid catalyzed halogenation, described in Section 20.2, proceeds through an enol.

❷ The acyl bromide is quite reactive and forms an enol more readily than the carboxylic acid.

❸ As can be seen by the resonance structure, the double bond of the enol is electron rich, so bromine adds rapidly by an electrophilic mechanism.

❹ Loss of a proton produces the α-brominated acyl bromide.



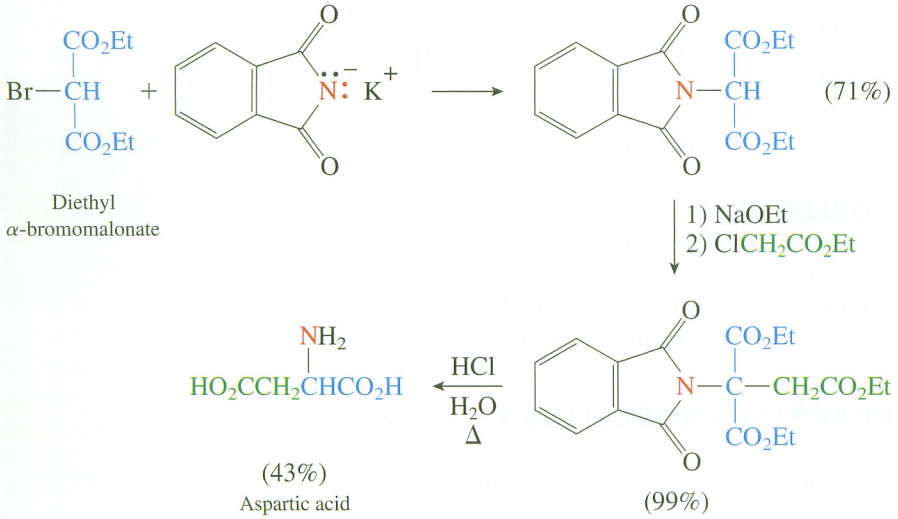
The acyl bromide can reenter the mechanism at step 2 and undergo enolization and bromination. Therefore, only a catalytic amount of PBr_3 is necessary.

❺ The Br of the acyl bromide exchanges with the OH of another molecule of the carboxylic acid to produce the α-brominated acid and an acyl bromide. This exchange reaction proceeds through an anhydride intermediate.

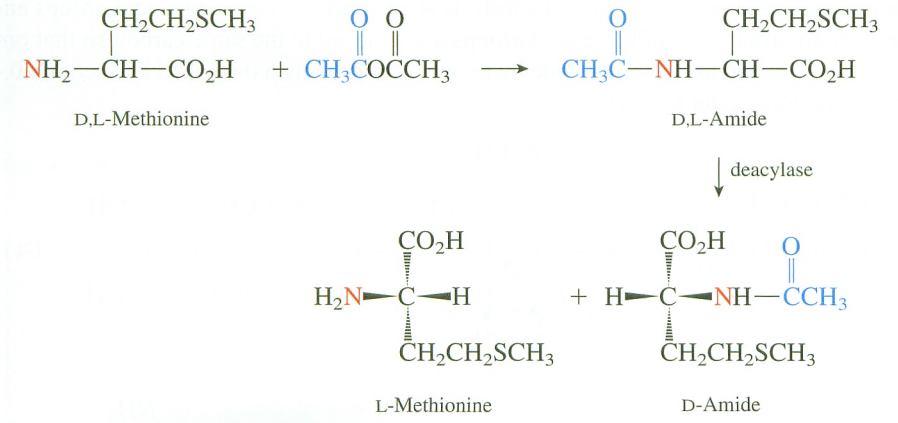
Figure 26.2

PARTIAL MECHANISM OF THE HELL-VOLHARD-ZELINSKY REACTION. Not all of the steps in the mechanism are shown.

A third amino acid synthesis begins with diethyl α-bromomalonate. First the Br is replaced by a protected amino group using the Gabriel synthesis (see Section 10.6). Then the side chain of the amino acid is added by an alkylation reaction that resembles the malonic ester synthesis (see Section 20.4). Hydrolysis of the ester and amide bonds followed by decarboxylation of the diacid produces the amino acid. An example that shows the use of this method to prepare aspartic acid is shown in the following sequence:

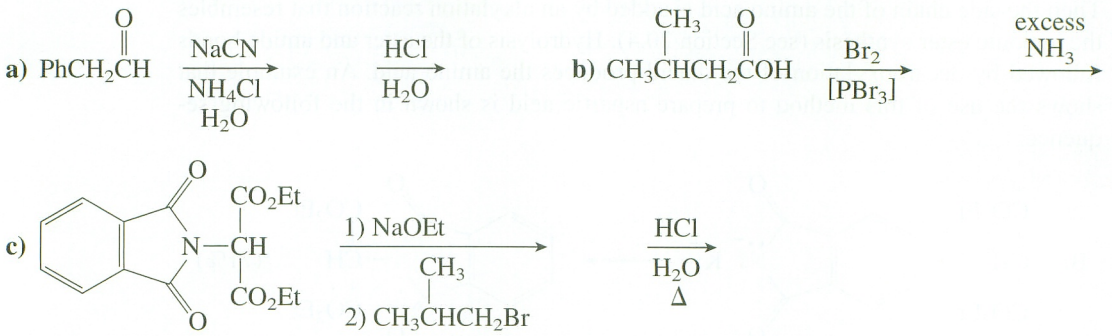


A drawback of all of these methods is that they produce racemic amino acids. If the product is to be used in place of a natural amino acid, it must first be resolved. This can be accomplished by the traditional method of preparing and separating diastereomeric salts. Alternatively, nature's help can be enlisted through the use of enzymes. In one method the racemic amino acid is converted to its amide by reaction with acetic anhydride. The racemic amide is then treated with a deacylase enzyme. This enzyme catalyzes the hydrolysis of the amide back to the amino acid. However, the enzyme reacts only with the amide of the naturally occurring L-amino acid. The L-amino acid is easily separated from the unhydrolyzed D-amide. The following equation illustrates the use of this process to resolve methionine:



PROBLEM 26.10

Show the products of these reactions:



PROBLEM 26.11

Show syntheses of these amino acids:

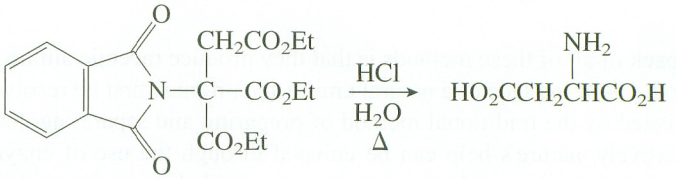
a) Leucine by the Strecker synthesis

b) Phenylalanine by the Hell-Volhard-Zelinsky reaction

c) Tryptophan starting from diethyl α-bromomalonate

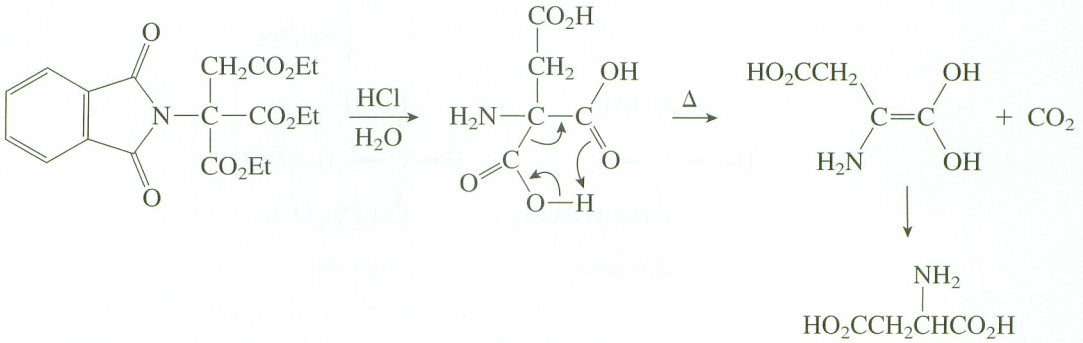
PRACTICE PROBLEM 26.2

Show representative steps in the mechanism for this reaction:



Solution

First all three ester bonds and both amide bonds are hydrolyzed to carboxylic acid groups by the aqueous acid. The mechanisms for these reactions are discussed in Section 19.5. The ester hydrolyses follow the exact reverse of the Fischer esterification mechanism shown in Figure 19.3, and the amide hydrolysis occurs by a very similar mechanism. The product of these hydrolysis steps has three carboxylic acid groups and one amino group. Two of these acid groups are attached to the same carbon so that one can be eliminated as carbon dioxide by the cyclic mechanism described in Section 20.4 for the malonic ester synthesis:

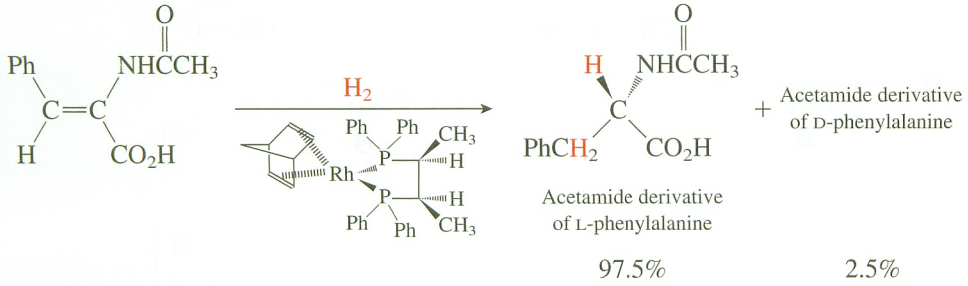


Focus On

Asymmetric Synthesis of Amino Acids

Even if the resolution of an amino acid is relatively easy, the synthesis of a racemic mixture when only one enantiomer is desired is wasteful, because half of the product cannot be used. Recently, considerable effort has been devoted to the development of methods that produce only the desired enantiomer by so-called asymmetric synthesis. As was discussed in Chapter 7, one enantiomer of a chiral product can be produced only in the presence of one enantiomer of another chiral compound. In some asymmetric syntheses a chiral reagent is employed. In others a compound called a chiral auxiliary is attached to the achiral starting material and used to induce the desired stereochemistry into the product. The chiral auxiliary is then removed and recycled.

An example of the use of a chiral reagent to accomplish an asymmetric synthesis of an amino acid is provided in the following equation:



The stereocenter at the α-carbon is introduced by catalytic hydrogenation. To selectively produce one enantiomer of the product, the acetamide of phenylalanine, a chiral catalyst is employed. Rather than using a metal surface as a catalyst, as is common for hydrogenations, a metal complex that is soluble in the reaction solvent is employed. In this particular case the catalyst is rhodium complexed to a bicyclic diene and a chiral phosphorus-containing ligand. Similar to the reaction on the surface of a metal, the reaction occurs by initial coordination of the alkene and hydrogen to the metal atom in place of the bicyclic diene. The chiral phosphorus ligand causes the hydrogen to be transferred to the alkene so as to produce a single enantiomer of the product. A variety of chiral catalysts have been developed, so one can often be found that will accomplish a particular asymmetric hydrogenation. In addition, the use of a chiral catalyst has the advantage that a full equivalent of the expensive chiral reagent is not needed and the catalyst can often be recovered. Another example of this process is provided in the Focus On box on page 449. You might recall that W. S. Knowles shared the 2001 Nobel Prize in chemistry for developing asymmetric hydrogenations like this.

The use of proline methyl ester as a chiral auxiliary in the asymmetric synthesis of alanine is shown on the following page. The idea is to start with 2-oxopropanoic acid (pyruvic acid), which has the correct carbon skeleton, and replace the oxygen on carbon 2 with an amino group and a hydrogen. This must be done in such a manner as to produce only the *S*-enantiomer of the amino acid, that is, L-alanine. This is accomplished by first attaching a chiral auxiliary, the methyl ester of L-proline, to the acid. In the critical step of the process, the catalytic hydrogenation, the chirality of the

Continued

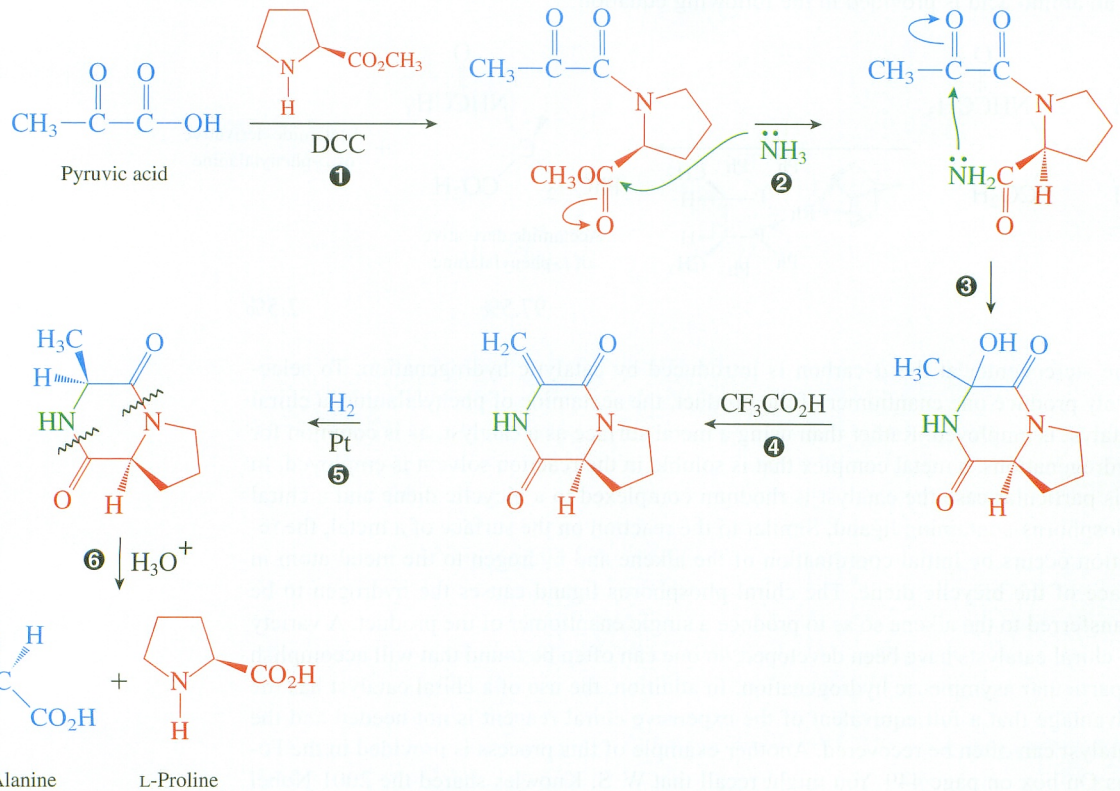
L-proline is used to induce the proper stereochemistry at the new stereocenter. To put this in terms used in Chapter 7, the α-carbon of proline has *S* stereochemistry. The new stereocenter generated in hydrogenation, which is the α-carbon of the alanine, could have either *R* or *S* stereochemistry. The potential products, with stereochemistries of *S,S* or *S,R*, are diastereomers. The hydrogenation occurs preferentially (greater than

The starting material for the synthesis of L-alanine is 2-oxopropanoic acid, also known as pyruvic acid. The carbonyl group at the α-carbon will be replaced with a H and a NH_2 so that only one enantiomer is formed at the new chirality center.

❶ In the first step, an amide is formed by the reaction of the acid with the methyl ester of L-proline, using DCC as the coupling agent (see Section 26.7).

❷ The product is then reacted with ammonia. The ammonia nucleophile attacks the carbonyl carbon of the ester group, resulting in the formation of an amide group. This product is not isolated but spontaneously proceeds to the next step.

❸ An intramolecular nucleophilic attack by the newly introduced nitrogen at the α-carbonyl carbon produces a six-membered ring.



❻ Acid-catalyzed hydrolysis of the two amide bonds produces L-alanine and regenerates the L-proline so that it can be used again.

❺ It is at this stage that the new stereocenter is introduced by catalytic hydrogenation of the double bond. The catalyst prefers to approach from the less hindered bottom side of the molecule, so a single stereoisomer of the product predominates. The yield of this step is quantitative, with more than 90% of the product having the stereochemistry shown.

❹ Trifluoroacetic acid causes the elimination of water to produce a CC double bond.

90%) at the less hindered bottom side of the molecule, so the hydrogen at the α-carbon of the alanine is added cis to the hydrogen at the α-carbon of the proline. This results in the formation of the *S*,*S*-enantiomer of the product. Hydrolysis produces the *S*-enantiomer of alanine, L-alanine, and regenerates the chiral auxiliary, L-proline, so that it can be used again.

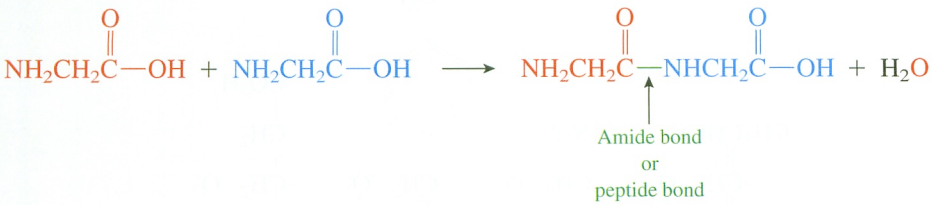
PROBLEM 26.12

What starting material would be used for the synthesis of L-phenylalanine by the method using proline methyl ester as a chiral auxiliary?

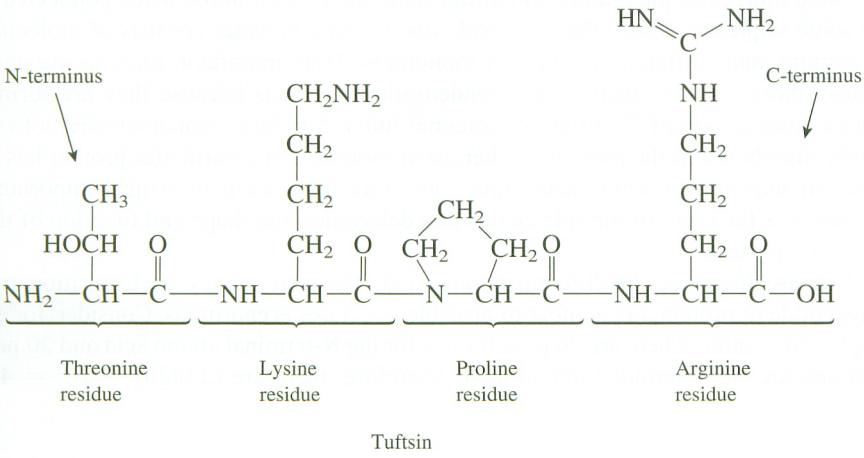
26.5 PEPTIDES AND PROTEINS

Because they contain two functional groups, amino acids can react to produce condensation polymers by forming amide bonds. These polymers are called peptides, polypeptides, or proteins. Although there is no universally accepted distinction, the term *protein* is usually reserved for naturally occurring polymers that contain a relatively large number of amino acid units and have molecular masses in the range of a few thousand or larger. The term *peptide* is used for smaller polymers.

As a simple example, the dipeptide formed by the reaction of two glycines has the following structure:



Biochemists say that the two amino acids are connected by a peptide bond, but, of course, the peptide bond is just an amide bond. A slightly more complex example, the phagocytosis-stimulating tetrapeptide known as tuftsin, derived from four amino acids, can be employed to illustrate some of the conventions that are used in writing the structures of polypeptides and proteins:



Tuftsin

By convention, peptides are written so that the end with the free amino group, called the N-terminus, is on the left and the end with the free carboxyl group, the C-terminus, is on the right. Because it takes considerable space to show the structure of even a small polypeptide like this one, it is common to represent the structures of peptides and proteins by using the three-letter abbreviation for each amino acid (see Table 26.1). Thus tuftsin, with a threonine N-terminal amino acid, followed by lysine, proline, and, finally, arginine as the C-terminal amino acid, is represented as

Thr-Lys-Pro-Arg

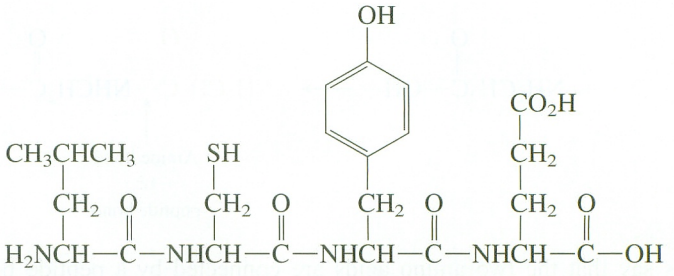
Note that the N-terminal amino acid is on the left and the C-terminal amino acid is on the right in this abbreviated representation also. (For very large polypeptides the one-letter codes for the amino acids are used to save even more space.)

PROBLEM 26.13

Draw the complete structure for Phe-Val-Asp.

PROBLEM 26.14

Identify the amino acids in this polypeptide and show its structure using the three-letter abbreviations for the amino acids:

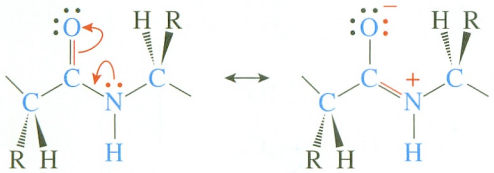


Let's compare proteins to the polymers that were discussed in Chapter 24. One difference is that all the molecules of a particular protein are identical; that is, they have the same molecular mass and contain the same number of amino acids connected in the same sequence. Recall that a typical condensation polymer consists of molecules containing many different numbers of monomers. More important, proteins are enormously more complex than simple condensation polymers because they are formed from a combination of 20 different monomer units. And these monomers are not randomly distributed in the protein. Rather, each molecule of a particular protein has an identical sequence of amino acid units. The exact sequence is of critical importance because it is the order of the side chains that determines the shape and function of that particular protein.

Because there are 20 different amino acids that can occupy each position in a polypeptide or protein, the number of possible structures is enormous. Consider, for example, a dipeptide. There are 20 possibilities for the N-terminal amino acid and 20 possibilities for the C-terminal amino acid. Therefore, there are $(20)(20) = 20^2 = 400$

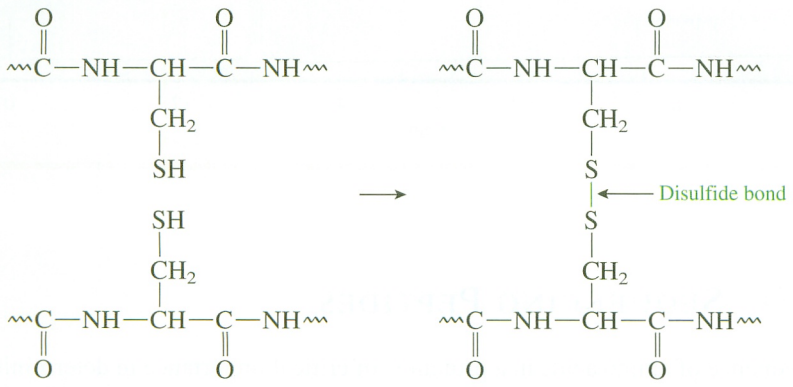
different dipeptides. The number of possibilities increases rapidly as the number of amino acids in the polymer increases. For a tripeptide there are $20^3 = 8000$ possibilities. And for a polypeptide that contains 100 amino acids (many proteins are considerably larger than this) there are $20^{100} = 1.27 \times 10^{130}$ possibilities. Such large numbers have little meaning for most of us, so let's try to put this number in perspective. It has been estimated that the total number of atoms in the universe is about 10^{80}. The number of possible polypeptides containing only 100 amino acids vastly exceeds the total number of atoms in the entire universe!

The geometry of the amide bond helps determine the overall shape of a peptide or protein. The nitrogen of an amide is sp^2 hybridized, so the electron pair on the nitrogen is in a p orbital that can overlap with the pi bond of the carbonyl group. The nitrogen is planar, and there is considerable double-bond character to the bond connecting it to the carbonyl carbon. In other words, the structure on the right makes an important contribution to the resonance hybrid for an amide:



This requires that the carbonyl carbon, the nitrogen, and the two atoms attached to each of them (the α-carbon and the oxygen bonded to the carbonyl carbon and the hydrogen and the other α-carbon bonded to the nitrogen) must all lie in the same plane. The most stable conformation has the bulky α-carbons in a trans relationship about the carbon–nitrogen partial double bond, as shown in the preceding structure.

Another important feature of proteins is due to the presence of cysteine amino acids in the polymer chain. The SH groups of two cysteine residues can react to form a disulfide bond as shown in the following equation:

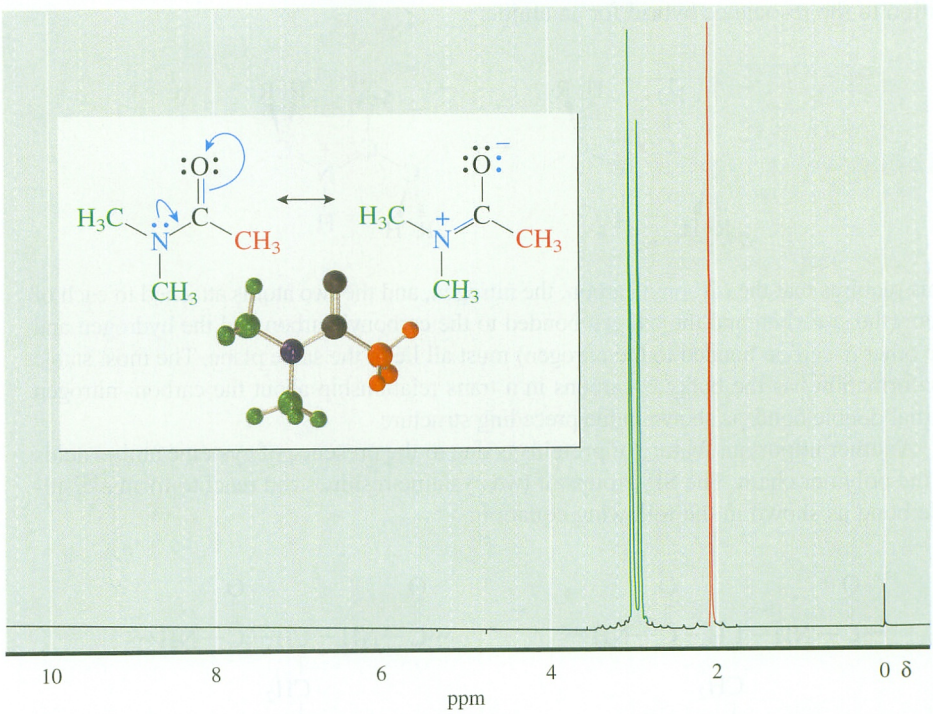


If the cysteines are part of different polypeptide chains, the resulting disulfide bond acts as a cross-link between the chains. If the cysteines are part of the same polypeptide chain, the large ring that is formed helps determine the overall shape of the peptide.

Focus On

NMR Spectra of Amides

Because the bond between the nitrogen and the carbonyl carbon of an amide has considerable double-bond character, rotation about this bond is relatively slow. This slow rotation is evident in the NMR spectra of many amides. For example, the two methyl groups on the nitrogen of *N*,*N*-dimethylacetamide appear at different chemical shifts. If rotation about the carbon–nitrogen bond were fast, the methyl groups would be equivalent and would appear at the same chemical shift.



26.6 Sequencing Peptides

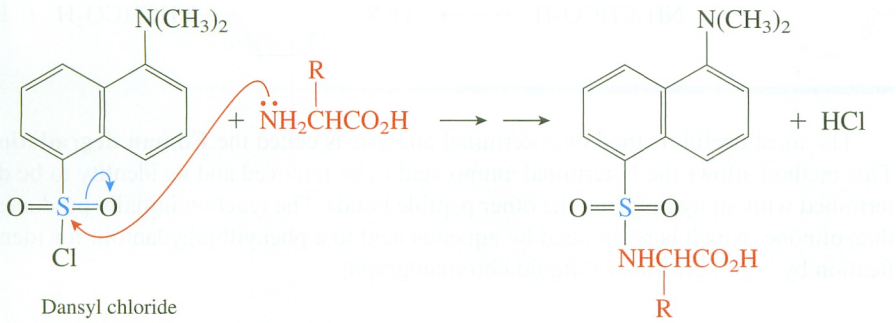
The sequence of amino acids in a protein is of critical importance in determining the function of that protein. Therefore, considerable effort has been devoted to the development of methods to determine amino acid sequences. The process usually begins with the determination of the relative numbers of each amino acid that are present in the protein. To accomplish this, a sample is completely hydrolyzed to its individual amino acid components by treatment with 6 *M* HCl at 100 to 120°C for

10 to 100 hours. These rather vigorous conditions are needed to completely hydrolyze the protein because amide bonds are rather unreactive. The amino acids are then separated by some type of chromatography, and the relative number of each is determined.

PROBLEM 26.15

Explain why asparagine and glutamine are never found when a peptide is completely hydrolyzed by using HCl, H_2O, and elevated temperatures. What amino acids are found in place of these?

The separation and detection process has been automated. In the original amino acid analyzer, developed by W. H. Stein and S. Moore, who were awarded the 1972 Nobel Prize in chemistry for determining the structure of the enzyme ribonuclease, the amino acids are separated by ion-exchange chromatography. They are then reacted with ninhydrin, and the resulting purple derivatives are detected by visible spectroscopy. In a more modern version the amino acids are reacted with dansyl chloride, and the resulting derivatives are separated by high-performance liquid chromatography. The dansyl group is highly fluorescent, so very small amounts of the dansylated amino acids can be detected. With a modern amino acid analyzer, the complete analysis of a hydrolyzed protein can be done in less than 1 hour. The method is sensitive enough to detect as little as 10^{-12} mol of an amino acid, so only a very small amount of the protein need be hydrolyzed.

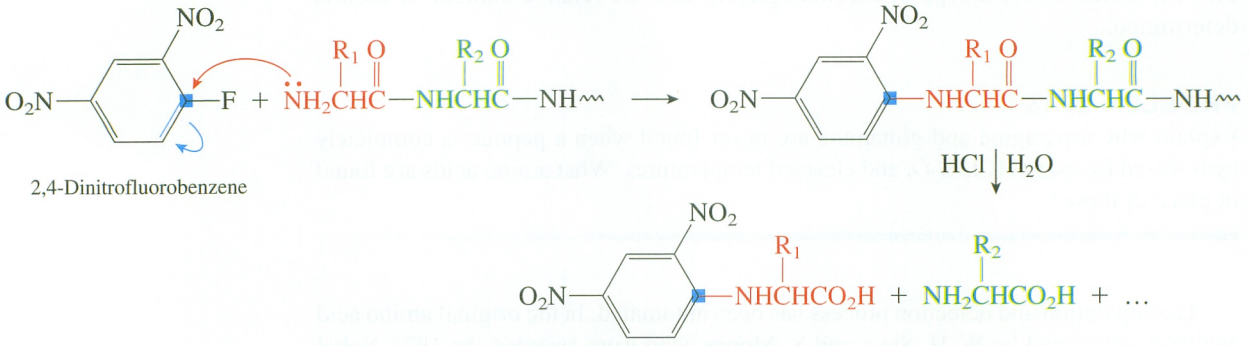


Dansyl chloride

Of course, the determination of the number of each kind of amino acid that is present is only a small part of the solution to the structure of a protein. The sequence of the amino acids must also be determined. This is accomplished by taking advantage of the fact that only the N-terminal amino acid has a free NH_2 group that is nucleophilic. All of the other nitrogens are part of amide groups and are not nucleophilic (unless the side chain contains an amino group, as is the case with lysine).

In one method the polypeptide is reacted with Sanger's reagent, 2,4-dinitrofluorobenzene (DNFB). The nucleophilic nitrogen of the N-terminal amino acid displaces the fluorine in a nucleophilic aromatic substitution reaction. (This reaction follows an addition–elimination mechanism; see Section 17.11.) The polypeptide is then hydrolyzed to its individual amino acid components. Because the bond between the nitro-

gen and the dinitrophenyl group is resistant to hydrolysis, the N-terminal amino acid is labeled and can easily be identified in the hydrolysis mixture:

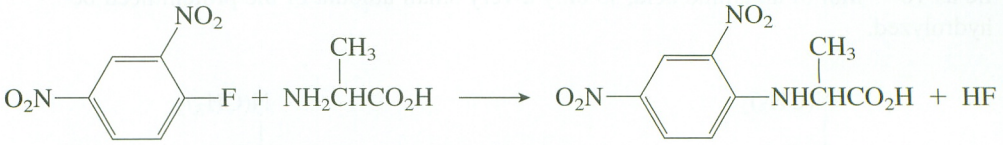


2,4-Dinitrofluorobenzene

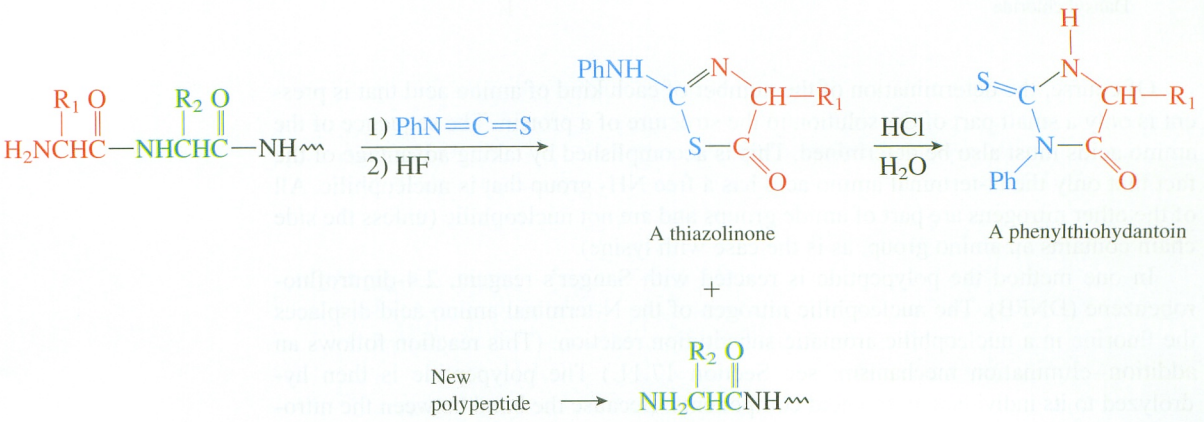
In another method the nitrogen of the N-terminal amino acid is reacted with dansyl chloride. The resulting sulfonamide bond is quite resistant to hydrolysis, so the N-terminal amino acid, labeled with the dansyl group, is readily determined after the peptide bonds have been hydrolyzed.

PROBLEM 26.16

Show all of the steps in the mechanism for the reaction of an amino acid with Sanger's reagent and explain why the nitro groups are necessary for the reaction to occur.



The most useful method of N-terminal analysis is called the **Edman degradation.** This method allows the N-terminal amino acid to be removed and its identity to be determined without hydrolyzing the other peptide bonds. The reaction initially produces a thiazolinone, which is rearranged by aqueous acid to a phenylthiohydantoin for identification by high-performance liquid chromatography:



The feature that makes the Edman degradation so useful is that the new polypeptide, with one fewer amino acid, can be isolated and submitted to the process again, allow-

ing identification of its N-terminal amino acid. It is possible to continue removing and identifying the N-terminal amino acid for 40 cycles or more before impurities due to incomplete reactions and side reactions build up to the extent that the identification of the last removed amino acid is uncertain. The Edman degradation procedure has also been automated, so it is now possible to sequence a polypeptide from the N-terminal end at the rate of about one amino acid residue per hour.

The critical feature of the Edman degradation is that it allows the N-terminal amino acid to be removed without cleaving any of the other peptide bonds. Let's see how this occurs. The mechanism of the reaction is shown in Figure 26.3. First the nucleophilic nitrogen of the N-terminal amino acid attacks the electrophilic carbon of phenyl isothiocyanate. When anhydrous HF is added in the next step, the sulfur of the thiourea acts as an intramolecular nucleophile and attacks the carbonyl carbon of the closest peptide bond. It is the intramolecular nature of this step and the formation of a five-membered ring that result in the selective cleavage of only the N-terminal amino acid. The mechanism for this part of the reaction is very similar to that for acid-catalyzed hydrolysis of an amide (see Section 19.5). However, because no water is present, only the sulfur is available to act as a nucleophile. The sulfur is ideally positioned for intramolecular attack at the carbonyl carbon of the N-terminal amino acid, so only this amide bond is broken.

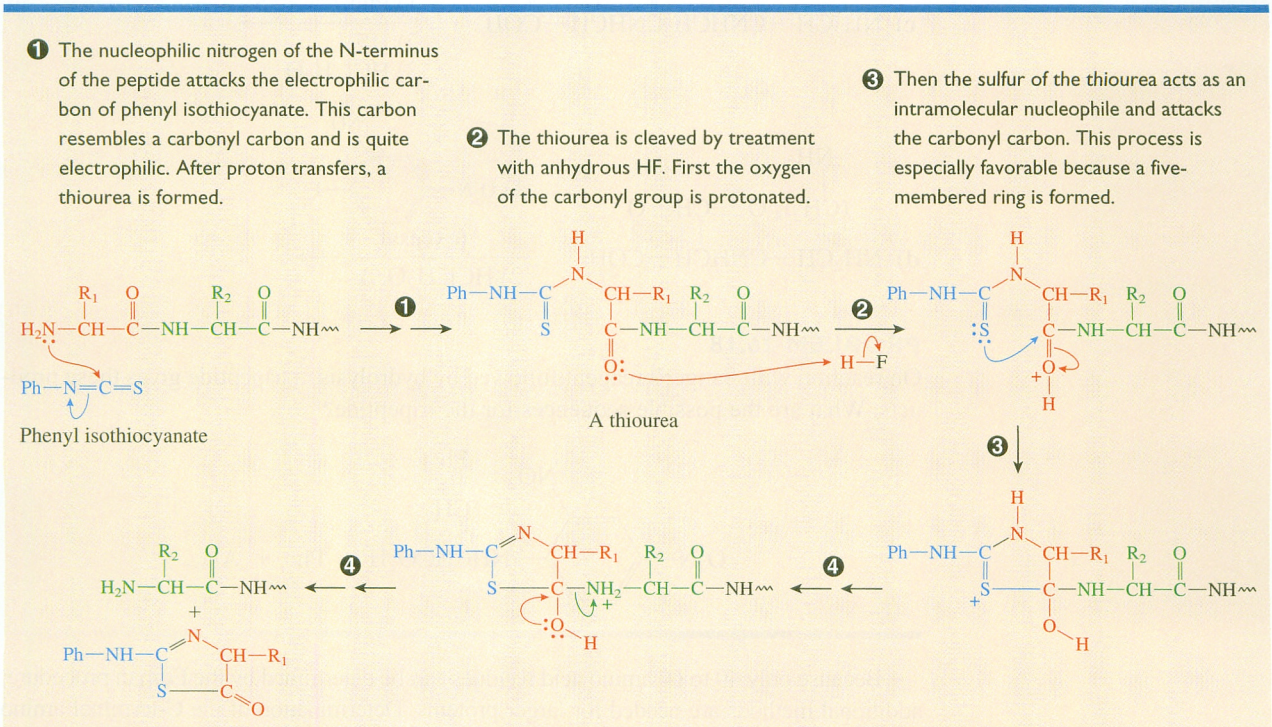
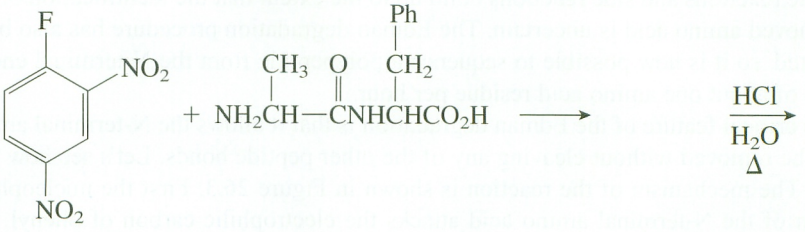

1 The nucleophilic nitrogen of the N-terminus of the peptide attacks the electrophilic carbon of phenyl isothiocyanate. This carbon resembles a carbonyl carbon and is quite electrophilic. After proton transfers, a thiourea is formed.
2 The thiourea is cleaved by treatment with anhydrous HF. First the oxygen of the carbonyl group is protonated.
3 Then the sulfur of the thiourea acts as an intramolecular nucleophile and attacks the carbonyl carbon. This process is especially favorable because a five-membered ring is formed.
Phenyl isothiocyanate
A thiourea
4 After several proton transfers, the nitrogen of the amide leaves. Overall, this mechanism is similar to acid-catalyzed amide hydrolysis (see Chapter 19), but the nucleophile is sulfur rather than water. No water is present, and the sulfur nucleophile can reach only the closest carbonyl group, so only the N-terminal amino acid is cleaved.

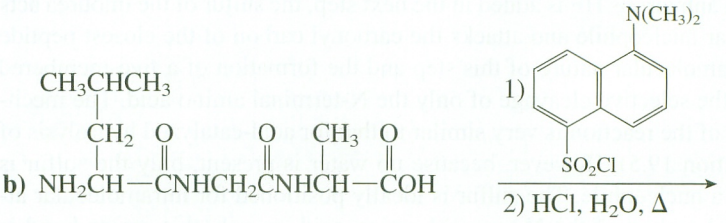
Figure 26.3

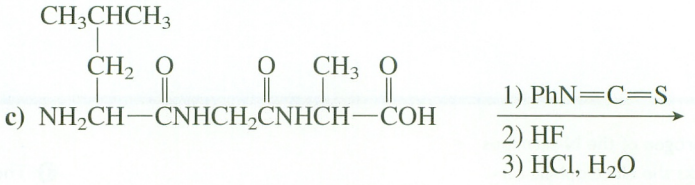
PARTIAL MECHANISM OF THE EDMAN DEGRADATION.

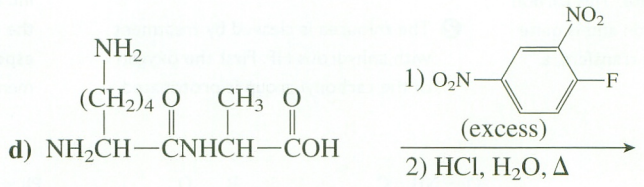
PROBLEM 26.17

Show the products of these reactions:

a)


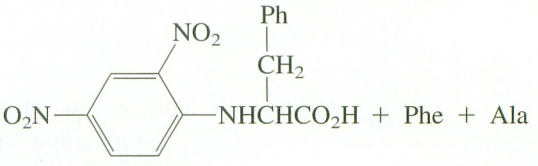
b)


c)


d)


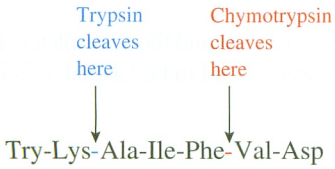
PROBLEM 26.18

On reaction with Sanger's reagent followed by hydrolysis, a tripeptide gives these products. What are the possible sequences for the tripeptide?



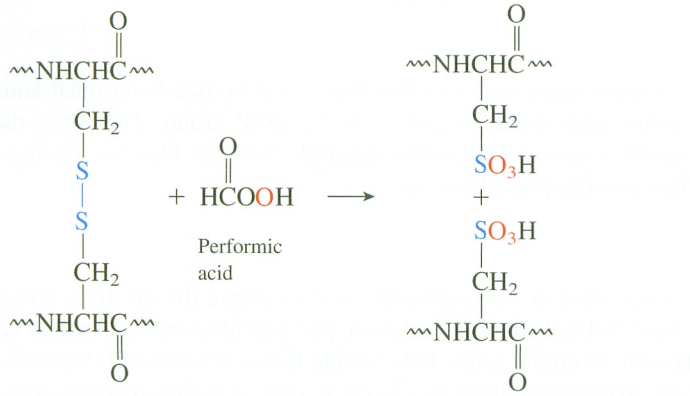
Because only 40 to 60 amino acid residues can be determined by the Edman procedure, additional methods are needed for larger proteins. Determination of the C-terminal amino acid can be accomplished by treating the protein with carboxypeptidase. This enzyme selectively catalyzes the hydrolysis of the C-terminal amino acid. After the first amino acid has been removed, the enzyme begins to cleave the second amino acid, and so forth. By following the rates at which the amino acids appear, it is possible to determine the first few amino acids at the C-terminal end of the protein by employing this enzyme. However, because the enzyme hydrolyzes different peptide bonds at different rates, it is possible to identify only a few amino acids before the reaction mixture becomes too complex.

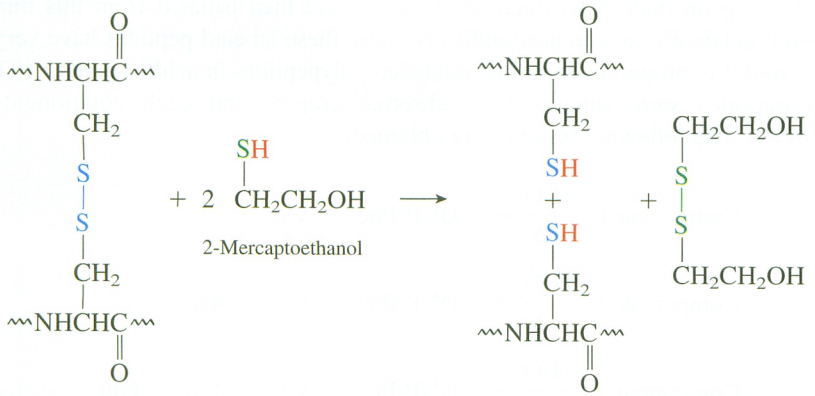
Most proteins are too large to be completely sequenced by using both N-terminal and C-terminal analysis. In such cases it is necessary to cleave the protein into smaller fragments that can be individually sequenced by using the preceding methods. Although random hydrolytic cleavage can be used, it is more common to use enzymes that cleave the protein in specific positions. For example, the enzyme trypsin cleaves peptides after amino acids with positively charged side chains—that is, after the basic amino acids lysine and arginine. Chymotrypsin cleaves after tyrosine, tryptophan, and phenylalanine, amino acids with aromatic side chains:



Several different cleavages must be done so that the fragments overlap and the amino acid sequence of one can be used to determine how others are connected.

A final problem is to determine which cysteines are connected by disulfide links. Usually, the disulfide bonds are broken by either oxidative cleavage with performic acid or reductive cleavage with 2-mercaptoethanol before initial sequencing is undertaken:



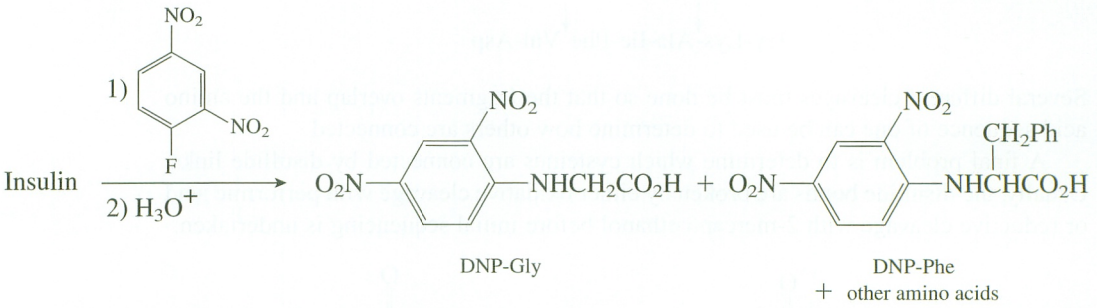


After the sequence of the individual chains has been determined, a sample of the protein is partially hydrolyzed without cleavage of the disulfide bonds, and the fragments containing the intact disulfide bonds are sequenced.

The first protein to have its sequence determined by these methods was insulin. This was accomplished in 1953 by the research group headed by Frederick Sanger. Sanger won the 1958 Nobel Prize in chemistry for directing this work. (He also shared the 1980 Nobel Prize for his contributions to DNA sequencing, so he is one of the few individuals to have won two Nobel Prizes.) Sanger chose insulin because it is readily available and is relatively small (51 amino acids). To help see the approach used to solve a problem of this type, let's examine a few of the many experiments that were used.

Experiment 1

Insulin was reacted with Sanger's reagent and then completely hydrolyzed. Two amino acids, a glycine and a phenylalanine, were found to be labeled with the 2,4-dinitrophenyl group.



This experiment indicates that there must be two N-terminal amino acids because two amino acids were labeled with the DNP group. Therefore, there must be two polypeptide chains, connected by disulfide linkages. One has glycine as its N-terminus and the other has phenylalanine.

Experiment 2

Insulin was reacted with performic acid to cleave the disulfide bridges. Two polypeptides were isolated from this reaction, polypeptide A with 21 amino acids and polypeptide B with 30 amino acids. Polypeptide B was reacted with Sanger's reagent and then partially hydrolyzed. This produced a very complex mixture containing individual amino acids, dipeptides, tripeptides, and so on. All of the peptides that had the dinitrophenyl group on their N-terminal amino acid were then isolated from this mixture. (This was relatively easy to accomplish because these labeled peptides have very different solubility properties from the unlabeled polypeptides in acidic solution.) The labeled peptides were separated by chromatography, and each component was hydrolyzed. The following results were obtained:

$$\text{Component 1} \xrightarrow[\Delta]{H_3O^+} \text{DNP-Phe} + \text{Val}$$

$$\text{Component 2} \xrightarrow[\Delta]{H_3O^+} \text{DNP-Phe} + \text{Val} + \text{Asp}$$

$$\text{Component 3} \xrightarrow[\Delta]{H_3O^+} \text{DNP-Phe} + \text{Val} + \text{Asp} + \text{Glu}$$

This experiment shows that Phe must be the N-terminal amino acid of peptide B because it is labeled with the DNP group. From hydrolysis of component 1, Val must be attached to Phe; from hydrolysis of component 2, Asp must be attached to Val; and from hydrolysis of component 3, Glu must be attached to Asp. The sequence of the first four amino acids of peptide B, starting from the N-terminus, is Phe-Val-Asp-Glu.

Experiment 3

It was known that there were two Cys residues in polypeptide B. All of the Cys-containing peptides from the hydrolysis mixture were isolated. This was accomplished by taking advantage of the strongly acidic sulfonic acid side chains of the oxidized cysteine residues. These peptides were separated by chromatography. Then each was reacted with Sanger's reagent and hydrolyzed, and its individual amino acids were identified. The following components were identified:

$$\text{Component 1} \xrightarrow[2)\ H_3O^+,\ \Delta]{1)\ \text{DNFB}} \text{DNP-Val} + \text{Cys}$$

$$\text{Component 2} \xrightarrow[2)\ H_3O^+,\ \Delta]{1)\ \text{DNFB}} \text{DNP-Leu} + \text{Cys}$$

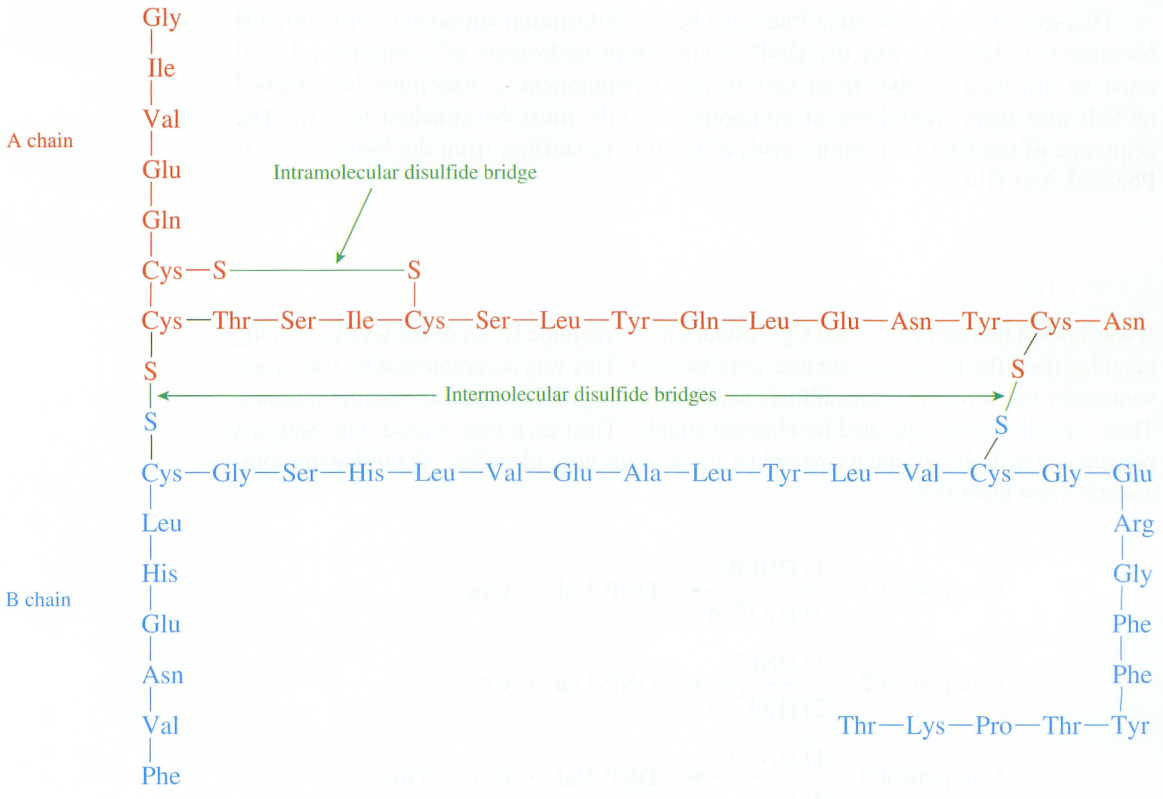
$$\text{Component 3} \xrightarrow[2)\ H_3O^+,\ \Delta]{1)\ \text{DNFB}} \text{DNP-Val} + \text{Cys} + \text{Gly}$$

$$\text{Component 4} \xrightarrow[2)\ H_3O^+,\ \Delta]{1)\ \text{DNFB}} \text{DNP-Leu} + \text{Cys} + \text{Gly}$$

$$\text{Component 5} \xrightarrow[2)\ H_3O^+,\ \Delta]{1)\ \text{DNFB}} \text{DNP-Leu} + \text{Cys} + \text{Val}$$

Component 1 indicates Val-Cys, and component 2 indicates Leu-Cys. (Remember that there are two Cys residues.) Component 3 indicates Val-Cys-Gly because the two Cys residues have Val and Leu attached to the amino end, so Val-Gly-Cys is not possible. Component 4 indicates Leu-Cys-Gly for similar reasons. Component 5 indicates Leu-Val-Cys because Val cannot be attached to the carboxyl end of Cys owing to the information obtained from components 3 and 4. Therefore, two small parts of the sequence are Leu-Val-Cys-Gly and Leu-Cys-Gly.

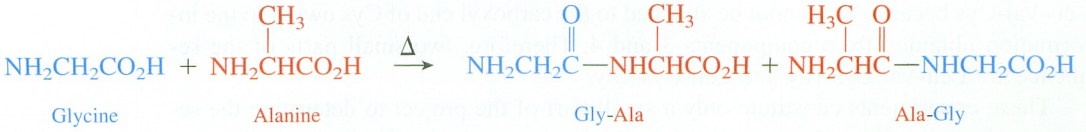
These experiments constitute only a small part of the project to determine the sequence of insulin. Overall, a large number of scientists worked under Sanger's direction on this project, which took 10 years and 100 g of protein to accomplish. Today, the sequence of such a simple protein could be determined by one experienced technician in only a few days using an automated sequencer employing the Edman method and would require only a few micrograms of sample! Or the sequence might be determined by methods employing mass spectrometry, or it might be read directly from the sequence of the DNA that codes for the protein.



Insulin

26.7 Laboratory Synthesis of Peptides

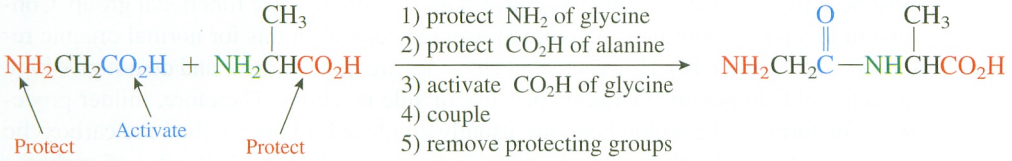
Synthesis of polypeptides in the laboratory is much more complicated than the synthesis of the polyamides described in Chapter 24 because more than one monomer must be used and the order of their attachment must be carefully controlled. For example, suppose we want to prepare the dipeptide Gly-Ala. We cannot just heat a mixture of glycine and alanine. Although this would produce some of the desired dipeptide, it would also form a host of other products, including other dipeptides and tripeptides, as shown in the following equation:



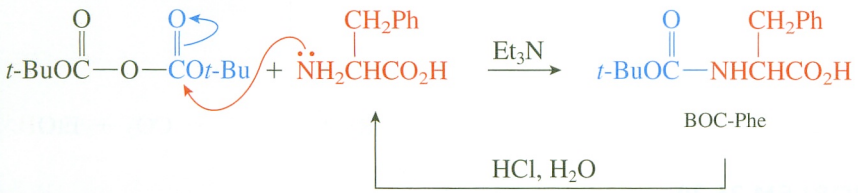
+ Gly-Gly + Ala-Ala + Gly-Ala-Gly + Gly-Ala-Ala + ...

The problem with this reaction is that there are too many reactive functional groups: two amino groups and two carboxylic acid groups. To prepare the desired dipeptide in a controlled manner, it is necessary to restrict the reactivity of the compounds so that only the amino group of alanine and the carboxylic acid group of glycine are available to react. As usual, this is accomplished by the use of protecting groups on the functional groups in which reaction is not desired. The synthesis of the dipeptide is performed by first protecting the

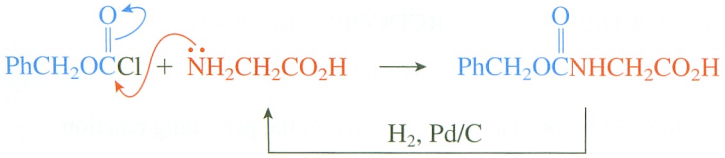
amino group of glycine and the carboxylic acid group of alanine. To form the amide bond, the carboxylic acid group of the protected glycine is activated by conversion to a more reactive derivative (see Section 19.6). Then the protected alanine is added. In the final step the protecting groups must be removed without breaking the newly formed amide bond.



Recall that an amino group is commonly converted to an amide for protection (see Section 23.4). However, there must be some special feature of the protecting group that allows it to be removed without also cleaving the amide bonds of the peptide. Although a large number of different protecting groups has been developed, the most common ones are the *t*-butoxycarbonyl (BOC) group and the benzyloxycarbonyl group. The BOC group is added by the use of an anhydride and is readily removed by treatment with dilute acid:

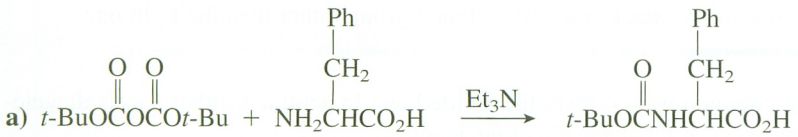


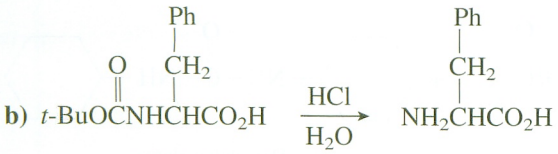
The benzyloxycarbonyl group is added by using an acyl chloride and can be removed by hydrogenolysis:

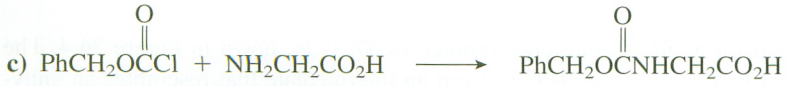


PROBLEM 26.19

Show all of the steps in the mechanisms for these reactions:

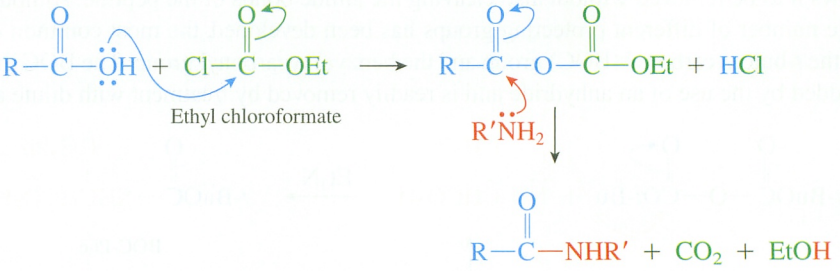






The carboxylic acid group is usually protected as an ester. Because an ester is considerably more reactive than an amide, it is possible to remove this protecting group by hydrolysis without cleaving the amide bond of the peptide.

For a carboxylic acid and an amine to form an amide, the carboxylic acid usually must be activated; that is, it must be converted to a more reactive functional group. Conversion to an acyl chloride is a common way to accomplish this for normal organic reactions (see Chapter 19). However, acyl chlorides are quite reactive and do not give high enough yields in peptide synthesis because of side reactions. Therefore, milder procedures for forming the amide bond are usually employed. In one method the carboxylic acid is reacted with ethyl chloroformate (a half acyl chloride, half ester of carbonic acid) to produce an anhydride. Treatment of this anhydride with an amine results in the formation of an amide:



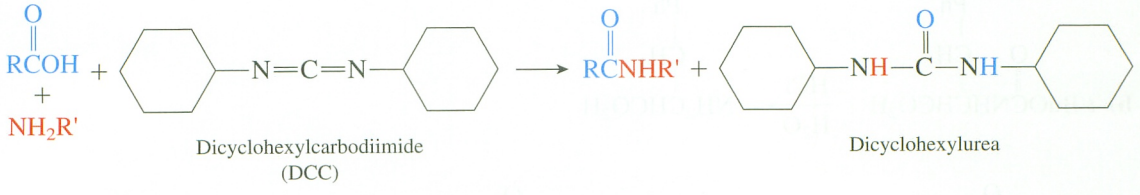
PROBLEM 26.20

Show all of the steps in the mechanisms for these reactions:

Explain why the second product is not formed in the preceding reaction.

Explain why the amine attacks the left carbonyl group rather than the right one.

The most common way to form the amide bond in peptide synthesis uses dicyclohexylcarbodiimide (DCC) as the coupling agent:



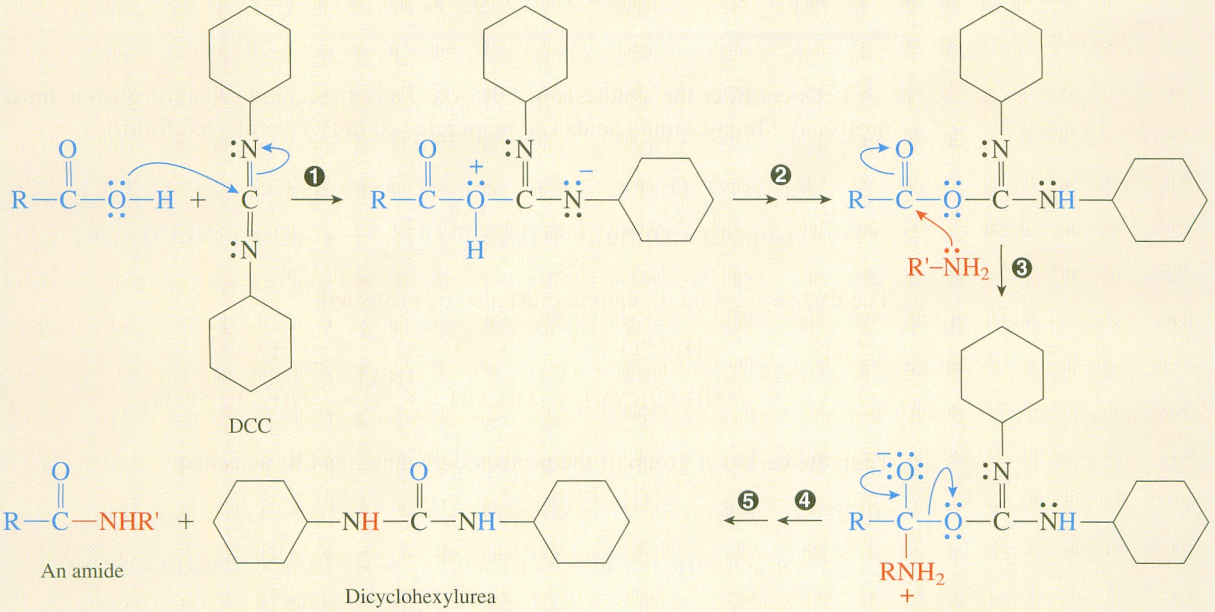
The mechanism for amide formation promoted by DCC is shown in Figure 26.4. The carboxylic acid first reacts with DCC to form an intermediate that resembles an anhydride, but with a carbon–nitrogen double bond in place of one of the carbonyl groups.

❶ The carbon of DCC is quite electrophilic. The oxygen of the carboxylic acid acts as a nucleophile and adds to the CN double bond in a reaction that is very similar to the addition of nucleophiles to the carbonyl group of an aldehyde or ketone (see Chapter 18).

❷ Proton transfers complete the addition reaction.

This compound resembles an anhydride, but with a CN double bond rather than a CO double bond.

❸ Just like the reaction with an anhydride, the amine nucleophile attacks the carbonyl carbon.



❺ Proton transfers result in the formation of the amide and dicyclohexylurea.

❹ The oxygen of the "anhydride" then leaves.

Active Figure 26.4

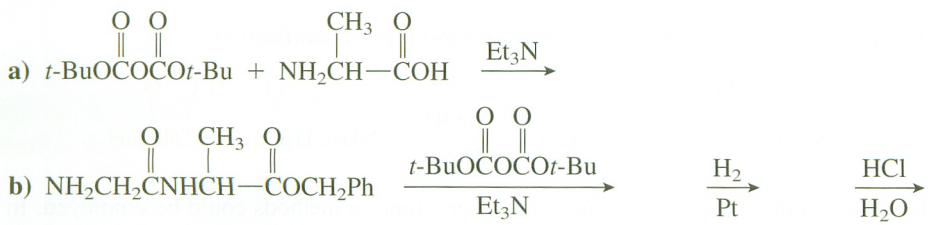
ORGANIC Chemistry Now™

MECHANISM OF AMIDE FORMATION USING DICYCLOHEXYLCARBODIIMIDE. Test yourself on the concepts in this figure at **OrganicChemistryNow.**

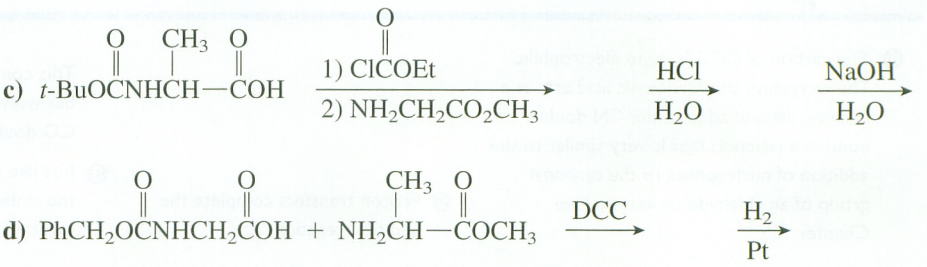
When this pseudo-anhydride is reacted with an amine, an amide is produced along with the by-product, dicyclohexylurea. Many other coupling reagents and protecting groups have been developed, but the ones shown here are the most common and illustrate the principles involved.

PROBLEM 26.21

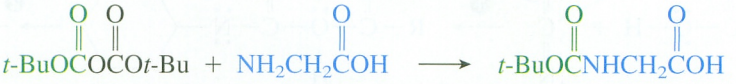
Show the products of these reactions:



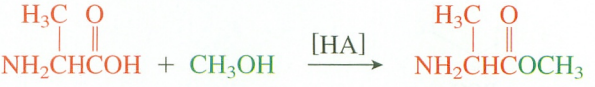
ORGANIC Chemistry Now™
Click *Coached Tutorial Problems* for more practice with **Reactions Used in Synthesis of Peptides.**



Let's consider the synthesis of Gly-Ala. First, the amino group of glycine must be protected. (Today, amino acids can be purchased in BOC protected form.)



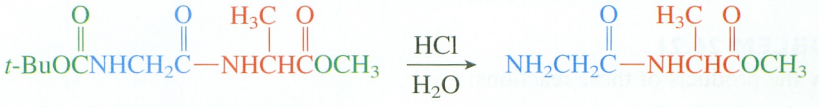
The carboxyl group of alanine must also be protected:



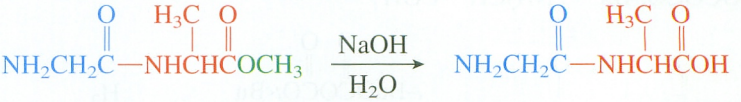
Then the carboxyl group of the protected glycine must be activated:

The next step is the formation of the amide bond by reaction of the activated, N-protected glycine with the carboxyl-protected alanine:

To complete the synthesis, it is necessary only to remove the protecting groups. The BOC group is removed by treatment with aqueous acid. This reaction occurs under very mild conditions that do not also hydrolyze the ester group. Note that this dipeptide, deprotected at the N-terminal, can be reacted with another N-protected, carboxyl-activated amino acid to produce a tripeptide. These steps can be repeated to produce a tetrapeptide, and so on.



Finally, the carboxyl protecting group is removed by saponification:



Of course, other protecting groups and other coupling methods could be employed. In addition, things are a little more complicated because many of the amino acids have

functional groups in their side chains that must also be protected. However, the general ideas of the process are provided in the preceding procedure.

PROBLEM 26.22
Show a synthesis of Ala-Phe-Gly-Gly starting from the individual amino acids.

As is apparent from this example, the synthesis of a polypeptide requires numerous steps. At each step, the product must be isolated, and if you have worked in an organic chemistry laboratory, you are certainly aware of how much time and energy are required to isolate and purify a product. This makes a polypeptide synthesis quite tedious. In addition, the mechanical losses that occur in each isolation step contribute to lower yields for the overall process. Motivated by these problems, R. B. Merrifield developed a method, called *solid phase synthesis,* that makes the preparation of a polypeptide much easier. Merrifield was awarded the 1984 Nobel Prize in chemistry for this work. Let's see how it works.

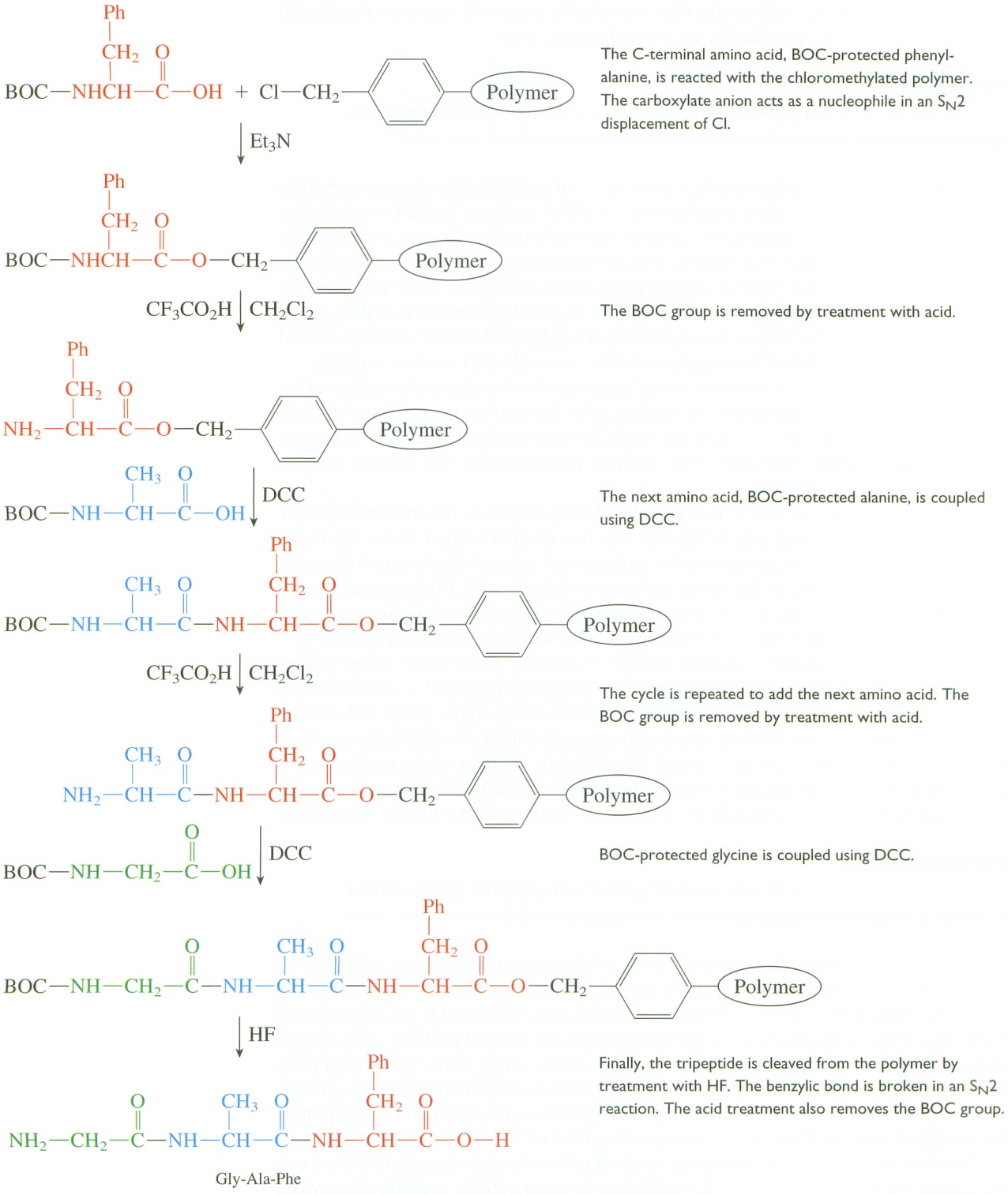
Merrifield's idea was to attach the initial amino acid molecules to the surface of insoluble polymer beads. To isolate the product after the next amino acid has been attached, it is necessary only to collect the beads by filtration and wash them to remove any remaining reagent. Not only is the isolation procedure fast and simple, but mechanical losses are minimized.

The insoluble polymer that is used for solid phase syntheses is the chloromethylated, divinylbenzene cross-linked polystyrene that was described in Section 24.10. The beads of this material have $ClCH_2$ groups bonded to the phenyl groups that are part of the polymer chains on the surface of the beads. As we saw in Section 24.10, functional groups on the surface of such beads exhibit normal chemical reactions. To synthesize a polypeptide, the carboxyl group of the C-terminal amino acid, with its amino group protected, is first bonded to the polymer by an ester linkage. The amino group of this amino acid is deprotected. Then the next amino acid, with its amine group protected, is attached using DCC to promote coupling. This cycle of deprotection of the amino group followed by coupling with another N-protected amino acid is repeated until all of the amino acids of the desired polypeptide have been added. At each step, isolation of the product is accomplished by simply collecting the beads by filtration. Finally, the peptide is cleaved from the polymer bead. An example of a solid phase synthesis is provided in Figure 26.5.

PROBLEM 26.23
Show a synthesis of Ala-Phe-Gly-Gly using the Merrifield solid phase method.

As mentioned earlier, the ease of isolation of the product is the major advantage of the solid phase method. However, this is also a disadvantage because the product, the growing polypeptide attached to the polymer, is never purified until it is finally cleaved from the polymer. Each time a coupling reaction does not proceed in 100% yield, a small amount of the final product will be missing one of the amino acids. These impurities build up with the number of steps in the synthesis and become more and more difficult to remove as the polypeptide becomes larger. For this reason the yield of each step must be as high as possible. Current methods provide yields of 99.5% or better in each step.

Another advantage of the solid phase method is that the steps are repetitive and can be automated. An automatic peptide synthesizer is commercially available. It consists of a flask with a fritted-glass filter at the bottom along with bottles containing solutions of BOC-protected amino acids and the other necessary reagents and wash solutions. The polymer with the first amino acid attached is placed in the flask, and the order of



The C-terminal amino acid, BOC-protected phenylalanine, is reacted with the chloromethylated polymer. The carboxylate anion acts as a nucleophile in an S_N2 displacement of Cl.

The BOC group is removed by treatment with acid.

The next amino acid, BOC-protected alanine, is coupled using DCC.

The cycle is repeated to add the next amino acid. The BOC group is removed by treatment with acid.

BOC-protected glycine is coupled using DCC.

Finally, the tripeptide is cleaved from the polymer by treatment with HF. The benzylic bond is broken in an S_N2 reaction. The acid treatment also removes the BOC group.

Figure 26.5

THE MERRIFIELD SOLID PHASE SYNTHESIS OF GLY-ALA-PHE.

the amino acids in the desired polypeptide is programmed into the machine. The synthesizer then adds the next amino acid and DCC, reacts the mixture for the appropriate period of time, and removes the reagents by filtration. After the polymer beads have been washed, the next reagent is added and the process continues. The nonapeptide hormone, bradykinin, has been synthesized in 85% yield in 27 hours using such a device. As another example, bovine pancreatic enzyme ribonuclease A, with 124 amino acids, was synthesized in 17% yield in 6 weeks. After purification this material showed 78% of the bioactivity of the natural enzyme. However, this is an example in which the synthesis went particularly well. More commonly, large polypeptides like this contain many impurities that cannot readily be removed and therefore have a bioactivity that is much less than the natural protein. Solid phase synthesis is best for smaller polypeptides, up to about 50 amino acid residues. For larger peptides it is usually best to synthesize fragments of 50 amino acids or less and then couple these by solution phase reactions after they have been purified.

26.8 Protein Structure

The properties of a protein depend primarily on its three-dimensional structure. The sequence of amino acids in the polypeptide chain is termed its **primary structure.** Its **secondary structure** is the shape of the backbone polypeptide chain. Remember that each amide group is planar, but the chain can have various conformations about the bond between the α-carbon and the nitrogen. The **tertiary structure** is the overall three-dimensional shape of the protein, including the conformations of the side chains.

The primary structure of the protein—that is, the sequence of its amino acids—determines its secondary and tertiary structure. Each protein usually has one shape that has the lowest energy. Individual molecules of the protein naturally and reproducibly fold into this lowest-energy conformation. The interactions that determine the preferred shape include hydrogen bonding involving the hydrogens on the amide nitrogens and the carbonyl groups, disulfide bonds, and interactions of the side chains with other side chains and the aqueous environment of the protein. The protein usually folds so that charged side chains—those of aspartic acid, glutamic acid, lysine, arginine, and histidine—are on the surface of the protein, so they can be solvated by water. Polar side chains also prefer the surface, where they are exposed to the aqueous environment, whereas nonpolar side chains, with their hydrophobic groups, prefer to pack together in the interior of the protein to avoid contact with water.

Two common elements of secondary structure that comprise about 50% of the structure of an average protein are the α-helix and the β-pleated sheet. In an α-helix the polypeptide chain assumes a helical shape that is held in place by hydrogen bonding between the NH group of one amino acid residue and the carbonyl oxygen four residues removed. There are 3.6 amino acid residues per turn of the helix. A simplified picture of an α-helix is shown in Figure 26.6.

In β-pleated sheets the hydrogen bonding occurs between parts of the polypeptide chain separated by larger distances. In one version, the antiparallel β-pleated sheet, the parts of the chain in the sheet run in opposite directions; that is, the N-terminal to C-terminal directions are opposite, as shown in Figure 26.7. In a parallel β-pleated sheet, the N-terminal to C-terminal directions are the same. In either case, when viewed from the edge, the sheet takes on a pleated shape. In an actual protein, α-helixes and β-pleated sheets, connected by loops and turns, pack together to give the overall three-dimensional shape of the protein.

Figure 26.6

PARTIAL STRUCTURE OF AN α-HELIX.

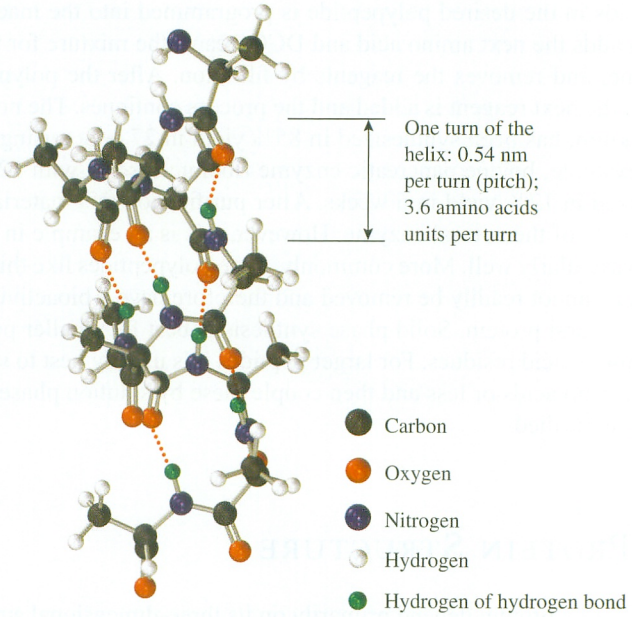
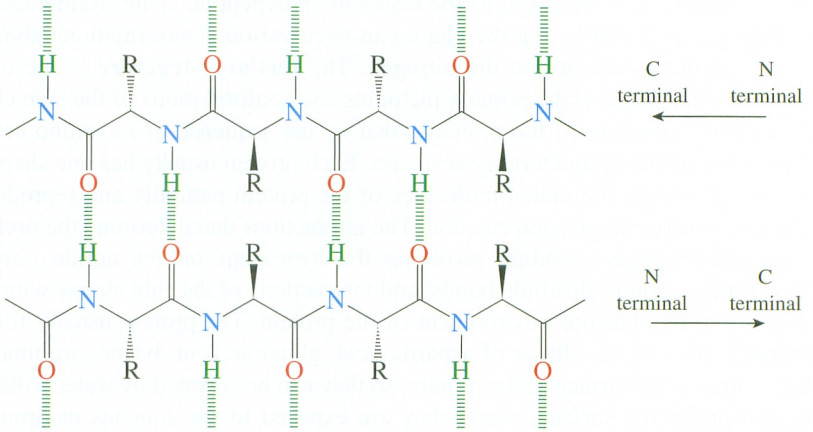

One turn of the helix: 0.54 nm per turn (pitch); 3.6 amino acids units per turn
Carbon
Oxygen
Nitrogen
Hydrogen
Hydrogen of hydrogen bond

Figure 26.7

PARTIAL STRUCTURE OF AN ANTIPARALLEL β-PLEATED SHEET.



ORGANIC Chemistry Now™
Click *Coached Tutorial Problems* to view an **α-Helix** or a **β-Sheet.**

26.9 ENZYMES

It was Emil Fischer who first proposed that an enzyme and its substrate (the compound undergoing the enzyme-catalyzed reaction) fit together like a lock and key. The substrate is tightly bound to the enzyme by hydrogen bonding, polar interactions, and hydrophobic interactions in a precise three-dimensional arrangement. In the active site of the enzyme–substrate complex, various functional groups of the side chains of the amino acids comprising the enzyme are held in proximity to the substrate. It is the proximity of these catalytic functional groups to the reactive site of the substrate that causes the rapid rate of catalysis. Essentially, the catalytic reaction resembles the intramolecular reactions that we have seen so many times to proceed at greatly enhanced rates.

As an example, let's consider the mechanism of action of the enzyme chymotrypsin. This enzyme is known as a serine protease because it catalyzes the hydrolysis of

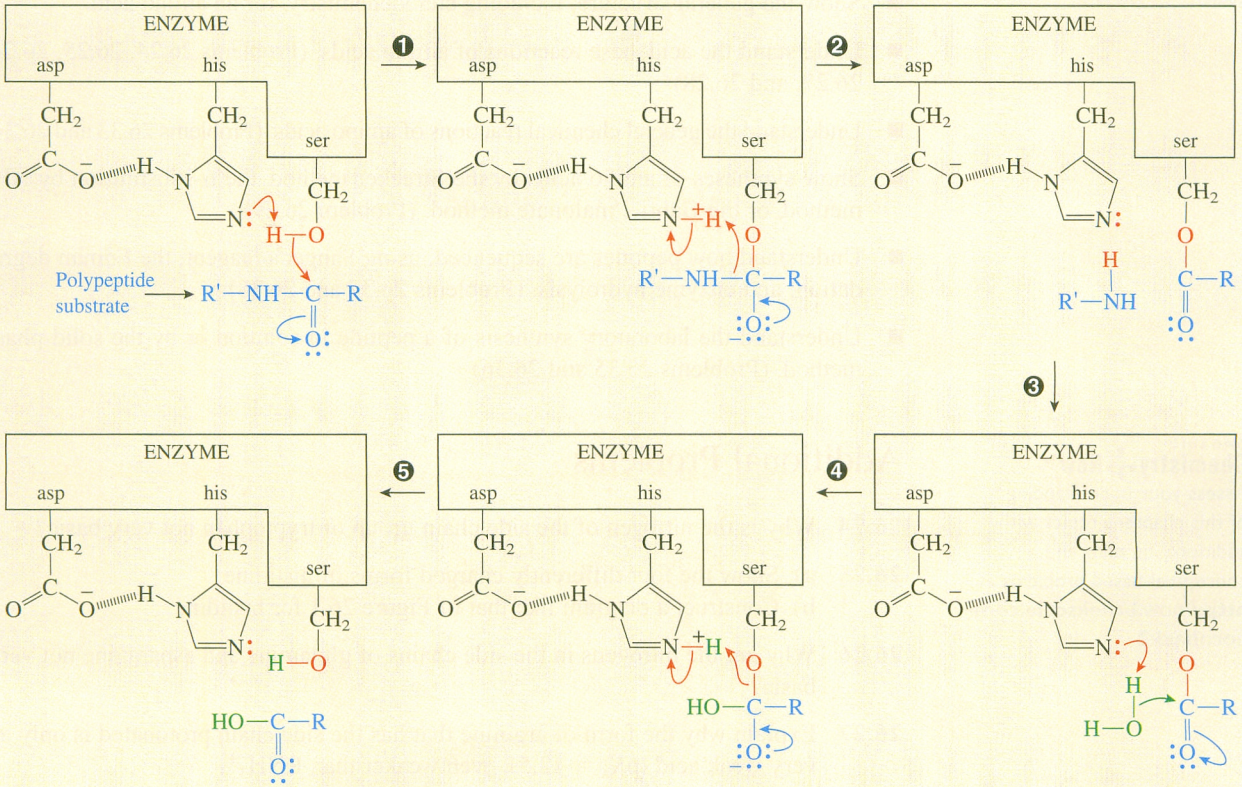
❶ The side chains of a serine, a histidine, and an aspartate residue are all important in the active site. The polypeptide substrate is held in proximity to the OH of the serine. The oxygen of this OH acts as a nucleophile and attacks the carbonyl carbon of the peptide. As this happens, the histidine removes the proton from the OH, making it a better nucleophile. And the aspartate makes the histidine more basic by hydrogen bonding to it.

This is the tetrahedral intermediate shown in the mechanism of Figure 19.1.

❷ As happens in that mechanism, the electrons on the negative oxygen displace the leaving group, the nitrogen in this case. The protonated histidine transfers its proton to the nitrogen to make it a better leaving group.

At this point, the amide bond has been broken and the amine part is free to diffuse out of the active site. The acyl group is bonded to the oxygen of the serine. To complete the catalytic cycle, this ester bond must be hydrolyzed and the OH of the serine must be regenerated.

❸ The amine diffuses away *and is replaced* by water.



The remaining steps are just the reverse of the preceding steps.

❹ The water acts as a nucleophile, attacking the carbonyl group. Again, the histidine removes the proton from the water as this reaction proceeds.

❺ The serine oxygen acts as a leaving group, generating the free carboxylic acid, which diffuses from the active site. The enzyme is now ready for another catalytic cycle.

Figure 26.8

MECHANISM OF PEPTIDE HYDROLYSIS CATALYZED BY CHYMOTRYPSIN.

polypeptides using the hydroxy group of a serine residue as a nucleophile. Its mechanism is outlined in Figure 26.8. First, the substrate, a polypeptide, is bound to the enzyme active site. (For simplicity this binding is not shown in the mechanism in Figure 26.8, but it is of critical importance because it makes the rest of the steps favorable by making them intramolecular. This binding also gives chymotrypsin its selectivity. Large hydrophobic groups, such as a phenyl group, are bound more tightly at the active site of chymotrypsin, so it preferentially cleaves peptides after amino acid residues with such

groups—that is, after phenylalanine, tyrosine, and tryptophan.) The hydrolysis mechanism follows the same steps as shown in Figure 19.1, with the oxygen of the hydroxy group of a serine acting as the nucleophile.

These last two sections have necessarily been brief, but they provide some of the general principles governing the structure of proteins and the action of enzymes. A textbook of biochemistry should be consulted for more details.

ORGANIC Chemistry Now™
Click *Mastery Goal Quiz* to test how well you have met these goals.

Review of Mastery Goals

After completing this chapter, you should be able to:

- Show the general structure, including stereochemistry, for an amino acid.
- Understand the acid–base reactions of amino acids. (Problems 26.24, 26.25, 26.26, 26.27, and 26.28)
- Understand the general chemical reactions of amino acids. (Problems 26.33 and 26.34)
- Show syntheses of amino acids by the Strecker method, the α-substitution by NH_3 method, or the Gabriel/malonate method. (Problem 26.29)
- Understand how peptides are sequenced, using Sanger's reagent, the Edman degradation, and enzyme hydrolysis. (Problems 26.37 and 26.38)
- Understand the laboratory synthesis of a peptide in solution or by the solid phase method. (Problems 26.35 and 26.36)

ORGANIC Chemistry Now™
Assess your understanding of this chapter's topics with additional quizzing and conceptual-based problems at **http://now.brookscole.com/hornback2**

Additional Problems

26.24 Why is the nitrogen of the side chain group of tryptophan not very basic?

26.25 **a)** Show the four differently charged forms of histidine.
b) Construct a diagram like that of Figure 26.1 for histidine.

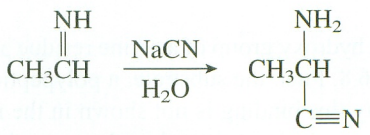
26.26 Why are the nitrogens in the side chains of glutamine and asparagine not very basic?

26.27 Explain why the form of arginine that has the side chain protonated is only a very weak acid ($pK_a = 12.5$), even weaker than RNH_3^+.

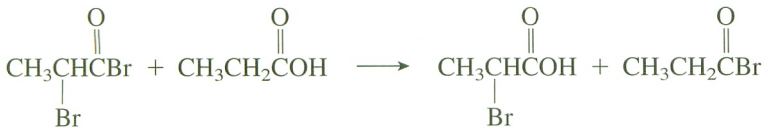
26.28 Explain which nitrogen of the side chain ring of histidine is protonated in the monocationic form.

26.29 Show syntheses of isoleucine by the Strecker method, the α-substitution by NH_3 method, and the Gabriel/malonate method.

26.30 Show a mechanism for the second part of the Strecker synthesis:

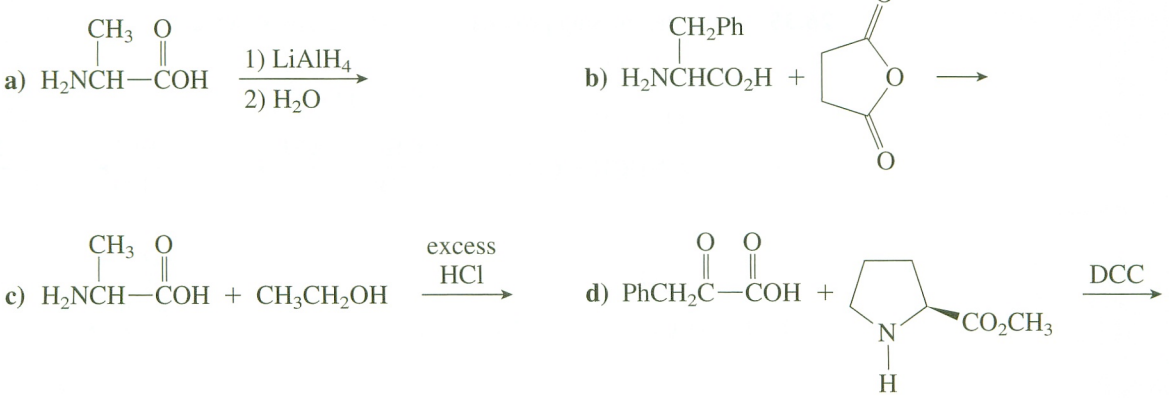


26.31 Show all of the steps in the mechanism for this reaction:

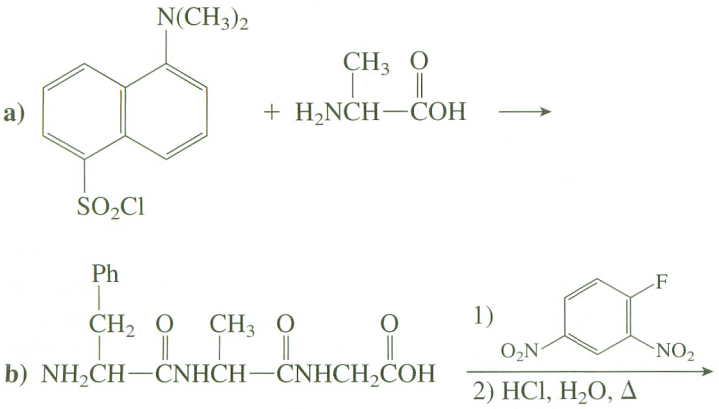


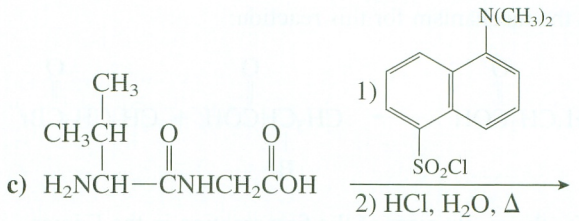
26.32 Show all of the steps for the mechanism of the first reaction in the Edman degradation:

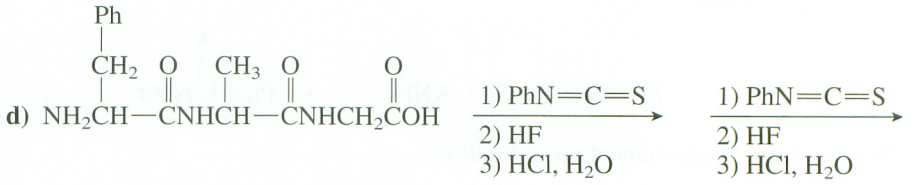
26.33 Show the products of these reactions:



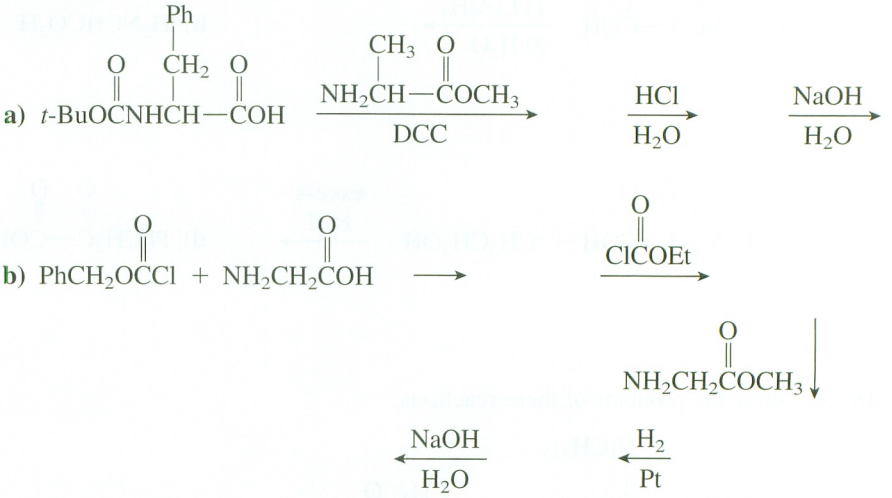
26.34 Show the products of these reactions:







26.35 Show the missing products in these reaction schemes:



26.36 Show a synthesis of Phe-Gly-Ala-Leu
a) by the solution phase method.
b) by the solid phase method.

26.37 An unknown pentapeptide is treated with Sanger's reagent and then hydrolyzed with aqueous HCl to produce Val, Gly, Leu, Ala, and Phe labeled with the dinitrophenyl group. After partial hydrolysis of the pentapeptide, three components were isolated from the complex reaction mixture. Each of these components was hydrolyzed with aqueous HCl to give the following results: component 1 produced Val and Gly; component 2 produced Phe, Leu, and Ala; and component 3 produced Gly and Leu. Deduce the structure of the pentapeptide.

26.38 Bradykinin is a nonapeptide that causes severe pain. It is formed in humans in response to stimuli such as a wasp sting. Complete hydrolysis of bradykinin produces 3 Pro, 2 Phe, 2 Arg, 1 Gly, and 1 Ser. After reaction with Sanger's reagent and complete hydrolysis, an Arg is labeled with the dinitrophenyl group. Reaction with carboxypeptidase first produces Arg, followed by Phe. Partial hydrolysis gives a complex mixture from which five components can be isolated. Treatment of each of these components with Sanger's reagent followed by hydrolysis gives the following results:

Component 1 gives 2 Pro and Arg labeled with a dinitrophenyl group;
Component 2 gives Arg, Phe, and Pro labeled with a dinitrophenyl group;
Component 3 gives Gly and Pro labeled with a dinitrophenyl group;
Component 4 gives Ser, Phe, and Gly labeled with a dinitrophenyl group;
Component 5 gives Ser and Phe labeled with a dinitrophenyl group.

Deduce the sequence of the amino acids of bradykinin.

Do you need a live tutor for homework problems? Access vMentor at Organic ChemistryNow at **http://now.brookscole.com/hornback2** for one-on-one tutoring from a chemistry expert.